Sean M. Slovney. Usage Trends in a Digital Library: A Case Study of Iupac.org. A Master's paper for the M.S. in I.S. degree. November, 2004. 93 pages. Advisor: Deborah Barreau.

The International Union of Practical and Applied Chemistry (IUPAC) is a digital library that is host to numerous collections and journals on chemistry. In response to increasing demand over the last few years, they want to improve the Iupac.org website. To do so, they must first determine how the site is being used, so that they may make the appropriate changes that would best improve the digital library. Therefore, the goal of this study is to analyze web usage logs and survey users so that future designs of iupac.org will adequately address the needs of its target audience.

Headings:

Digital Libraries

Surveys

Web Analytics

USAGE TRENDS IN A DIGITAL LIBRARY: A CASE STUDY OF IUPAC.ORG

by
Sean M. Slovney

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

November, 2004

Approved by:

Deborah Barreau

Table of Contents

List of Tables

List of Figures

I. Introduction

The International Union of Practical and Applied Chemistry (IUPAC), is a chemical digital library that is home to collections of specialized journals and articles on chemistry. Over the past nine months, a web programmer was hired to increase the functionality and usability of the site by adding a search engine, forums for users, and creating and maintaining website statistics. Now that most of these elements have been implemented, the Board of directors at Iupac.org is faced with one question: **What should be done to Iupac.org in order to make the site better for its intended audience of chemistry enthusiasts and professionals?** In order to answer this question, we first need to know how the site is being used, so that we will know what needs to be changed and what does not.

Directly asking the users about Iupac.org is the best way to determine exactly how the site is being used, and the types of problems and issues they encounter when visiting the site. A Web-based survey was developed using PHPesp, an open source web survey tool that dynamically generates a Web-based survey that users can access and complete. Each of the survey responses is stored in an online database, identifiable only by its unique survey number, which is extremely helpful in ensuring the privacy and confidentiality of respondents. The methods of this study were modeled after the research methods of two site usage studies on digital libraries, one by Steve Jones, and another by Michael P DAlessandro, M.D. However, there is a distinct difference

between this study and the two studies mentioned. Both studies used surveys that specifically focused on aspects of the search engine, while the survey used in this study asked questions about the search engine as well as questions about other aspects of the site, including the collections used and the architecture of information.

The survey complemented the section of the study that examined site usage statistics provided by web server logs to look at what, why, and how people are doing things on the site itself. These statistics will help determine if there are any trends in things such as the average time that people stay on the site, what days/hours of week seem to generate the most traffic, and the most frequent search terms used in the search engine. The other difference between this study and the studies conducted in the past is the statistical package used to analyze the web server logs. A web metrics program called Urchin is used in this research to generate reports and graphs from the web server logs for IUPAC. Studies in the past have used custom analytical packages created specifically for use with their studies, while this one used a commercial statistical application.

Twenty-five participants, all adults over the age of 18 years of age, were asked to complete the survey online. When the search engine was initially created, IUPAC's board of directors asked 25 people within IUPAC to test it out and give feedback on its performance. The 25 participants are composed of some of the board of directors, members of IUPAC, and visitors who frequently use Iupac.org; essentially people who have a vested interest in the success of the site. Therefore, those same 25 individuals were asked to participate in the study. Among those 25, the exact gender breakdown was not known, but it is likely that the target participants included both males and females.

The participants were recruited through a mass email and the only inducement that was used to garner their participation was mentioning the potential improvements to the site that could come from their participation in the study. Because the survey was anonymous due to the way PHPesp records responses, there was no risk to the participants, and their responses were private and confidential.

II. Background

The first step in figuring out how to study and answer the research question is to do a literature review of previous research in the general area of interest relating to digital library usage. "Scholarly Communication and the Digital Library: Problems and Issues" by Steven P Harter and "Digital Libraries – Some Analog Issues" by Jacob D Vakkayil are articles that cover all facets involved in the creation and maintenance of digital libraries as well as touch on the issues and problems that are associated with them. "Bits and Bytes and Still a Lot of Paper: Astronomy Libraries and Librarians in the Age of Electronic Publishing", by Uta Grothkopf on the other hand talks about the advantages and disadvantages of digital library usage. It is essential to review this information before even considering how to make an existing digital library better for its potential users.

"Transaction Log Analysis of a Digital Library" by Steve Jones, Sally Jo Cunningham, Rodger McNab and Stefan Boddie, and "Evaluating Overall Usage of A Digital Health Sciences Library" by Michael P DAlessandro, M.D., Donna M. DAlessandro, M.D. Jeffrey R. Galvin, M.D., William E. Erkonen, M.D. are studies on digital library usage that document and analyze site usage. Each of the studies used web server transaction logs to generate reports on usage and surveys to gauge user satisfaction with search engine results. The methods used in these studies can be replicated to approach the task of redesigning particular aspects of the site to make them more usable.

Finally, given the fact the studies on digital library usage used web analytics software, it will be useful to research the methodology behind such applications. '*A Survey of Web Analytics*' by Dhyani & Bhowmick, although on the low side of the practical business use of web analytics, provide good insight as to how the web analytics tools work, which will be discussed in greater detail later on. There were also three other articles which focused more on the practical usage of web analytics, and the importance of businesses utilizing these metrics to help them reach return on investment, which are the following: 'Web Performance Analytics That Matter' by Keith Regan, '*Web Analytics That Matter*' by Susannah Patton, and '*E-Analytics: Business Analytics For the New Economy*', an executive summary by the web analytics firm, NetGenesis.

III. What is a Digital Library?

A digital library, also known as a data warehouse, electronic library, or virtual library, is a collection of digital representations of numerous types of media, such as documents, images and sounds that are stored in an information repository and are available either through a local computer network or anywhere via the Internet. They typically support the functions of a traditional library rather than replacing them by providing online methods of searching and browsing for content in an always organized and efficient manner. Naturally, providing materials in a digital format makes the content much easier to manage, store, search for and retrieve. Libraries in the private and public sector, as well as government agencies and educational institutions have realized this, and as a result digital library systems are being adopted at a rapid rate.

Advantages. In "Bits and Bytes and Still a Lot of Paper: Astronomy Libraries and Librarians in the Age of Electronic Publishing", Grothkopf mentions 5 advantages to offering materials digitally: powerful search capabilities, ease of navigation through materials, ease of updating/correcting materials, ease of referencing, and ease of availability. The first, powerful searching capabilities, makes it much easier to find what you are looking for. In most digital library systems, materials are searchable through retrieval mechanisms that run queries against the content in the digital library, and return matching results. Initially, this is much easier than searching through card catalogs or physically searching for the materials in a library. However, today most libraries have

computer based cataloging systems, so more than likely the average person would not have to physically search for materials, but you would still have to physically retrieve it. In a digital library setting, all a patron would have to do is click on the link to retrieve the document instantly. This also makes it much easier for the patron to find exactly what he or she is looking for when they are not sure of exactly what they want. Instead of physically retrieving each book and skimming through it for content, a digital library can be used to retrieve each book instantly without having to go anywhere, and skim through it at their leisure.

The second advantage is ease of navigation through content. With a digital library, patrons can navigate to specific sections of documents with ease, such as specific chapters, references, graphics or charts, as well as doing searches for specific words or phrases within the documents, making the decision process much easier. It is much more difficult, if not impossible to do the same thing with hard copies of the documents. And again, patrons can do this at their leisure whenever they have the time to do it and have access to the system.

Ease of updating or correcting a document is the next advantage. Using textbooks as an example, most people will agree that each new edition of a text is just basically an updated version of the previous edition, with no more than a few pages, at most an entire chapter that has been added or updated. New editions of textbooks are usually published once every other year, and in some extreme examples, as much as twice a year. From a librarian's perspective, sometimes, as far as space is concerned it would be much easier to keep only the most recent edition of a text, and discard all previous editions. But that usually does not happen and most editions of a text are kept, perhaps because in specific

situations it is more desirable for the library to keep multiple if not all editions of a text. However, digital libraries offer a more practical solution in a situation where space is an issue. Given the fact that the texts would be stored digitally, more than likely with each text being in portable data format (pdf) with each chapter being a separate PDF file, adding or updating the content would be as easy as literally opening the file and making the appropriate changes. If space issues were present, and the library could not store each separate edition of the text, it could just simply update the text as new information or findings were discovered. If space is not an issue, the library could easily store each edition of the text with the system.

Ease of referencing is the next advantage, and it ties in with the navigation features of a digital library. Each document stored with the system will list any references that were used during the creation of the document. If the reference is available with the system, it will be available for download or viewing. This feature greatly adds to the overall experience for patrons. With a physical library, if a patron wanted to know more about a topic after reading a book by following up on the references, he or she would have to physically search for it and if the library has it, physically retrieve it. But with a digital library system, all of an article's or book's references are essentially a click away.

Accessibility is the last and probably most appealing advantage of a digital library system. As mentioned earlier, all materials are available anytime of the day, or if it is a private system, whenever you have access to the system. In addition, there are no limits to how many people can view a document at once or for how long. Physical libraries usually have a few copies of each article or book, and usually allow people to borrow

them for a set period of time. With a digital library, this is not an issue, since numerous people can view the same document simultaneously for as long as they like. Also, electronic versions of journals are usually available before printed versions are published, and users can browse the contents tables of forthcoming issues, giving digital library patrons an added incentive to browse electronic versions of journals.

Disadvantages. Despite all of the advantages of digital libraries, there are just as many drawbacks. Grothkopf identifies reading and browsing issues, bandwidth issues, added expenses to the customer, package, or institution, and added expenses for the library are the four main disadvantages to using a digital library system. Issues such as these are normal when trying to offer content online over a distributed network and are unavoidable. These systems can provide a great service to patrons and enhance the overall quality and effectiveness of a library, but strategies on how to deal with the drawbacks must first be considered before even considering implementing a system.

Reading and browsing online content is the first drawback. When faced with reading such long articles, Grothkopf states that people will either print them out if they are not too long, or if the article is indeed too long to print out, they will read it in periods, taking breaks in between readings, making note where they left off so that they will know what part of the article/book to continue from. Even with today's technology, online articles/books are not suitable for reading without printing them out. Currently, there is practically no way around this, other than to possibly break large articles/books into multiple files, which would make it easier to print them out.

Bandwidth issues are the next drawback with using digital libraries, and again, it is something that is inevitable. Network/online resource usage and overall performance has a negative or inverse relationship - as the number of people who are simultaneously using the same resource increases, the level of overall performance will decrease. A classic example of this relationship is performance on an online class registration system. At 8 am when the system first comes online on the first day of registration, system performance is extremely sluggish due to the high volume of users attempting to simultaneously access the system at once. Digital Library systems have the same types of issues, and again, they are unavoidable. Possible ways around this could be to increase the amount of bandwidth that is normally available to the system, which could be accomplished by mirroring the content available on the system to other servers, reducing the overhead on the system or increasing the overall network throughput to the system.

Added expenses to the customer/package/institution is another drawback. As mentioned earlier, when people are faced with reading an online article that is long, they tend to print the articles out, rather than reading them through a computer monitor. Printing out such large articles/books will shift the printing costs from the publishers to the customers. In addition, more than likely over time some digital libraries will not offer materials for free because they will not receive enough funding to support the hardware and software infrastructure, and will start to charge patrons for subscriptions to access and even to print articles/books. As the costs for hosting this material increase as they have been doing for the last couple of years or so, digital libraries that do not have stable sources of funding will be forced to charge patrons for usage.

Criteria for Inclusion. Every university or institution that has a digital library has criteria for what materials are allowed to be included with the system, since it is not practical to include everything. This criteria usually reflects the interests of the institution that is running the digital library, and since interests vary from university to university and institution to institution, it is rare to find two digital libraries that have the exact same criteria. Rather than generally addressing some of the things that most institutions look for in material that will potentially be digitally archived, the Columbia University digital library will serve as an example. The Columbia University system is a model example for other universities and institutions to follow. Their criteria system, available at http://www.columbia.edu/cu/libraries/digital/criteria.html, is broken down into five main sections: value of material, demand of material, intellectual property rights of the material, preservation of the material, and technical feasibility of hosting material. Materials being considered for digital preservation must meet the requirements of each section before being included with the system.

The first aspect that is looked at is the overall value of the material that is being considered. Redundancy is one thing that digital libraries try to avoid, and one of the first things that will be looked at is whether or not anything like the material being considered is actually hosted by them. Another thing to look at is the importance of the material as it relates to the overall understanding of the topic that it is representing. For example, if a librarian is looking at adding material to a Renaissance era collection on the system. Materials on the artists and scientists that contributed to the Renaissance would add to the overall understanding of the Renaissance period. Materials being added to the system should also strengthen materials that are already present with the system. It may be

possible that the material that is being considered to be added may reinforce content on the system that was previously not sufficient to represent a specific topic or concept. And finally, materials that can enhance the image of the institution or university are very welcome, as many institutions take pride in hosting collections or material that is very rare and is unique to the collection.

The second aspect that is looked at is the overall demand of the material that is being considered. Essentially, will the benefits of having this material justify the effort and costs required to acquire and preserve the material? If an institution wanted to create a large collection on the artifacts of the Stone Age, it would certainly want to have content that accurately represents the period. So in this situation, materials relating to the Stone Age would be in high demand. Will the material satisfy the current audience of patrons that normally visit the system? A digital library that specialized in artifacts of the Civil War would not add materials relating to the Bronze Age because it would not satisfy the current audience of patrons. Will this material attract a new audience even if the current attendance rates are low? Being the only library that has exclusive access to content of high interest to most people, such as artifacts from the Titanic, would be a good example of material that would more than likely increase system attendance. Will the material bring about lucrative collaborations between other institutions? If the University of North Carolina (UNC) is considering creating a Civil War collection and the University of Virginia (UVA) and Duke University, which both have Civil War Collections, hear of this they may offer to extend their services to UNC in return for UNC's participation a collaborative effort to accurately represent the war. Opportunities such as this are things that many institutions look forward to and value.

Intellectual property rights representation for the material being considered is the next aspect that is looked at. Does the institution legally have the right to offer the material that is being considered? Can the institution accurately give credit to the individual(s) who created and/or own the intellectual rights to the material? Will the institution have to limit or restrict access to the material based on the intellectual property laws that apply to the material? Can the institution offer the materials in accordance to all applicable intellectual property laws, and will any special provisions have to be made to do so? All of these questions and many more will have to be considered and researched before making a decision on whether or not to include the material with the system.

The fourth aspect is preservation of the materials being considered. Can the material that is being considered be digitally preserved safely while not causing any damage to it? Some materials, especially old letters and photos tend to be extremely fragile, and any attempts to handle them without extreme care could cause permanent damage. However, there is an advantage to taking the risk if indeed the document is fragile, and it would reduce the overall handling of the material if the digital preservation was successful. In addition, the added benefit would be that patrons would now have access to materials that they would normally not have access to because of their fragility in the past. Will the preservation protect material that is at a high risk of theft or mutilation? For example, some libraries host extremely rare and rather delicate letters, like the Declaration of Independence. Surely one would not want such an important and rather priceless piece of history on display. If someone were to somehow steal it and bring it to the black market, it could easily be worth millions of dollars. However, by

digitally preserving the document, one could still give people access to the material without putting the material in danger of being stolen or mutilated.

The fifth and final aspect that is considered is the technical feasibility of being able to host the material. Is it technically possible with current technology to capture, present, store and host the material being considered? Will the materials display well in a new digital format? How long will the digital content last before needed a rescan or re-preservation, taking current technology and new technological advances on the horizon? All these questions and many more will be carefully researched and answered before making a final decision as to whether or not the material in question will be included with the system.

Once material is reviewed and meets all the required criteria to be included with the digital library, it is digitally preserved with the system. Depending on the system that is being used by the institution, the preservation process may be as simple as just placing the material on a scanner and scanning it, or as complicated as using advanced software and hardware to prepare and preserve it.

Methods of Preservation. Digital preservation or digital archiving is the process of ensuring the longevity of electronic documents through advances in technology. Digital libraries are composed of digitally preserved materials that are archived with the system and are accessible as long as the system is up and running. Over the years, as new advances in technology were discovered, it became much easier to preserve materials, and as a result new methods of archiving were born. Currently, although there are probably hundreds different ways to digitally preserve materials, chances are each one probably falls in one of the following four methodologies described

by Alison Bullock in “Preservation of Digital Information: Issues and Current Status”: migration, emulation, hybrid conversion, or preservation of legacy technology.

Migration is one of the most common ways to digitally preserve materials. It covers everything from copying, converting or transferring digital information from one generation of technology to another one, ensuring the longevity of that information. So for example, copying digital information from a medium that is becoming obsolete or physically deteriorating to a newer one, such as copying information that is on a floppy disk to a DVD is migration. Other examples could be converting documents from one format to another, such as ASCII format to PDF format, or moving documents from one operating platform to another, such as from Microsoft Windows to UNIX.

Emulation, as Bullock describes it is the process of creating new software that replicates the functions of older software or hardware in order to reproduce its performance. Although not a flawless replication of the original technology, it is enough to provide the intended features and functionality of the content as it was originally available with the original software or hardware. New advances in software and hardware do not necessarily mean backwards compatibility with older technology, and when trying to integrate the two together, librarians quickly realized that it may be necessary to keep legacy systems in order to keep the materials that were dependent on them. Emulation started to gain notoriety mainly as a virtual modeling simulator specializing in “what if” scenarios with different configurations of systems and hardware, but over the years it attracted the attention of the digital preservation community after they realized that it might be a possible solution to preserving digital materials that could not be preserved through general migration methods. Suppose a library has a situation

where it has a Windows based digital content viewing application, and after switching to a Linux based operation system, the application does not work anymore. Rather than keeping the system that has the original content viewing application, the library could use an emulator to simulate the functionality of the original application on the new system, saving the cost of maintaining more systems than needed, or re-preserving the content that depended on the older application.

Hybrid conversion is a method of essentially taking a document and preserving it in two or more forms of media. This is usually done when you have a document that cannot be accurately archived by preserving it in just one form of media. Although it is not practical to preserve every single document in a hybrid conversion, for some documents it is the only way to fully capture all features of the document. A classic example of this is creating both microfilm and digital copies of a document to reformat the paper originals of that document. The digitally preserved hypertext copy of the paper document enhances access and functionality of the original, and the microform copy acts as an archival surrogate to the original. However, hybrid conversion has quite a few drawbacks, with added redundancy to the system, something that most digital libraries strive to avoid.

Preservation of the legacy technology is probably the last resort for a digital library in the event that migration and emulation efforts attempted do not yield acceptable results. Although this method would preserve the software and hardware dependent content, it would be at the cost of preserving and maintaining software and hardware that is obsolete and was probably supposed to be removed when the newer hardware and software were implemented. As a result, in addition to paying for maintenance and

technical support for newer systems, librarians will have to pay the same if not more for the legacy system, something that was probably not taken into consideration during budget planning. Yearly budgets are far from unlimited, and the unexpected costs of maintaining a legacy system could prove devastating to a budget. However, if a library must have the material in its native format with original layout and functionality, keeping the legacy systems may be the only option, and therefore the cost of maintaining the systems would be justified.

Once all materials in a digital library have been digitally preserved, they are placed within the system, and are then available to be accessed by users. Digital library systems range from simple web pages that provide links to digital content, or vast systems like Science Direct, ACM Portal or MetaPress, which provide powerful search functions to help patrons.

Challenges. Providing materials online or over a network for patrons to use is no small accomplishment. Even after all of the careful planning and testing of a digital library system, there can always be issues that no matter how well the system is implemented. Issues with administration, storage, presentation, classification, and retrieval are the most common. However these issues are challenges to the system rather than disadvantages, because these aspects must be present in a system. Although these aspects are challenges, if handled correctly, they can be a plus with a digital library system.

Administration of access to the system is the first of many challenges faced by digital libraries. When hosting digital content, the administrators of the system must ensure that the content being provided for public access can be accessed by everyone

while personal and private collections have restricted access so that only a single individual or a select group of individuals has access to it. These measures must be taken in order to protect certain materials from unauthorized access, use or disclosure. Administrators must also protect the identities of all users, to ensure that users are comfortable browsing not just controversial material, but all materials.

Storage is another obstacle for a digital library system. Most systems use a database of some sort to manage the content, usually a Database Management System (DBMS). There are two types which are normally used in conjunction with digital libraries: relational database systems and object oriented database systems. Whatever system used, it must be capable of storing all the digital material for the digital library, which in many cases tends to be a lot of data, in a variety of formats while providing access to this material potentially 24 hours a day. However, the real issue here is not so much the ability to store every single thing on the system, but rather offering as much of it in quick a period as possible, or at least at an acceptable rate of transfer. Fortunately, there are ways to improve access performance, one of them being compression techniques.

Text only documents in formats such as HTML, SGML and ASCII can be stored using compression formats like ZIP and TAR and reduce the overall storage size of a document by as much as 60 percent. There are also audio, video and image compression standards such as JPEG, MPEG and MP3 which can also significantly reduce the overall size of a file. Using compression techniques to reduce overall file size will naturally mean smaller files, which will in turn mean shorter amounts of time to access these files, which on a broad scope will mean improved performance with the system. Another way

to improve system performance is to use hierarchical storage mechanisms (HSM). Basically, the way an HSM works is that all data stored on the system are spread across numerous hard drives, more than likely in a redundant arrays of inexpensive disks (RAID) configuration. Content that is more frequently used is stored on hard drives with faster access speeds to accommodate the more popular content that most of the user's access, thus improving the performance of the system. System performance is the number one priority with these systems, and because of these requirements, a lot of time and money is invested in these systems to ensure that performance is optimal and that they stay up and running.

Just as most things are judged solely on their appearance, presentation in a digital library can be the deciding factor as to whether or not a patron wants to continue using it. Using a graphical user interface (GUI) for navigating through the content is practically the norm now, due to society's dependence on using GUI's for just about anything computer related. But, designing an intuitive and effective interface is not easy, and requires a considerable amount of research and requirements gathering from the potential users of the system, namely digital library patrons. Heedlessly rushing a GUI for a system could prove to be disastrous, because a bad GUI could possibly deter patrons from using a system. In addition to a good GUI, digital library presentation systems must be flexible in display and output options for content. Not everyone will have the latest hardware and software to view the materials at the best settings technologically possible, and with that in mind considerations must be made to accommodate the possibility of such a situation. Offering various display options for content such as different

resolutions for images or different connection speeds for streamed video media are ideal examples of trying to accommodating all patrons that might visit the system.

Classification of content is the next challenge faced by digital library systems. All content that is collected by the system must be classified with related content into groups before being put into the system. Usually the content is arranged in groups that are intuitive to patrons, so that theoretically they can find what they are looking for. However, given the fact that individual perceptions of what should be grouped where does vary from patron to patron, there will always be users who will not agree with the classification schemes, and as a result issues with groupings usually turn out into endless topics of debate among patrons and librarians alike. In addition, classification can be a slow and extremely tedious process which can lead to exorbitant amounts of material waiting to be indexed.

Information retrieval, which in itself can be a challenge, is something that must also be looked into when offering content through a digital library. There are three ways digital materials are usually searched in a system: searching by subject, full text document searching, or metadata searching. Out of the three, metadata searching would arguably appear to be the most accurate and effective because of the way metadata is structured and describes content with its attribute fields. But, no matter how sophisticated a retrieval system is, it is useless if it cannot not provide the majority of users with the information that they seek. Therefore, to aid retrieval mechanisms with digital library systems, some institutions offer user profiles or agents for use with the system. These profiles are designed to aid the user in finding what he or she is looking for based on their interests, so that when they do searches on the system, the results are

filtered for preferences, which will hopefully return more meaningful and valuable results to the user.

IV. International Union of Pure & Applied Chemistry

Fig. 1. Iupac.org website, Iupac.org 25 October 2004: http://iupac.org

The International Union of Pure and Applied Chemistry (IUPAC) is a not-for-profit international organization that is dedicated to the advancement of the global understanding of chemistry and all related sciences throughout the world. Driven by the desire to create international standards for the field of chemistry, IUPAC was founded in 1919 by chemistry experts and enthusiasts in industry and academia alike. IUPAC has garnered a reputation as being a world leader in chemical nomenclature, chemical

terminology, and standardized methods for measurement for chemical and atomic materials. It is responsible for the following standards that are present in modern chemistry today:

- Standardization of the symbols and terminology in chemistry,
- Standardization of atomic weights,
- Standardization of physical constants,
- Editing tables of properties of matter,
- Standardization of methods of data analysis and presentation,
- Standardization of the formats of publications,
- Nomenclature of inorganic and organic chemistry, and
- Data exchange standards for computers and instruments

In the 85 years since it has been in existence, IUPAC has also strived to be the liaison between the research interests of industrial, public, private, and academic sectors through international meetings and conventions. Their members are composed of two associations of international bodies, the National Adhering Organization (NAO), and the Associate National Adhering Organization (ANAO).

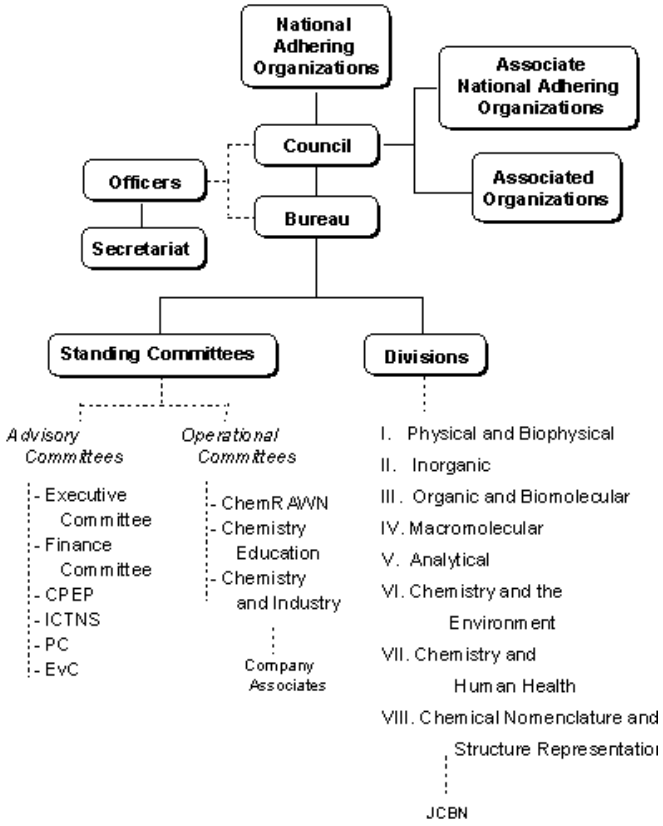


Fig. 2. IUPAC member chart, Iupac.org 25 October 2004: http://iupac.org

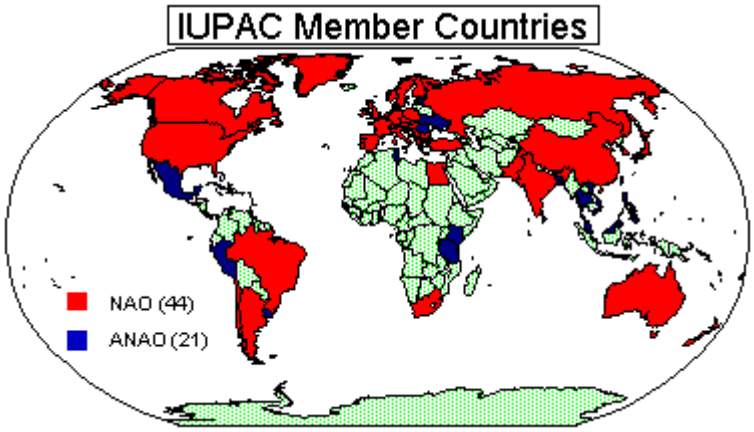


Fig. 3. IUPAC member countries map, Iupac.org 25 October 2004: http://iupac.org

The NAO has 45 countries who are members, and the ANAO has 20 member countries. Combined, there are 65 countries and over 1,000 chemists throughout the world that volunteer their scientific knowledge to Iupac, primarily through projects which are in any one of the eight divisions of research that IUPAC specializes in the following:

1. Physical and Biophysical Chemistry
2. Inorganic Chemistry
3. Organic and Biomolecular Chemistry
4. Macromolecular
5. Analytical Chemistry
6. Chemistry and the Environment
7. Chemistry and Human Health
8. Chemical Nomenclature and Structure Representation

Iupac.org, the official website for the organization, is the online digital library which represents the wealth of knowledge that Iupac has amassed over the years.

Infrastructure. As mentioned earlier, some digital library systems use custom or proprietary hardware and/or software to run their digital libraries. However, such is not the case with Iupac.org. The IUPAC needs are relatively small compared to corporations and institutions, and can easily provide the content without the need for such hardware or software. IUPAC's target audience is chemistry professionals and enthusiasts who are looking for scholarly publications and journals relating to chemistry.

To satisfy this audience, IUPAC has decided to provide unrestricted online access to chemistry related journals, books and reports.

The procedure for reviewing material for inclusion with the digital library tends to be a fairly simple process. All of the journals and reports are sponsored by IUPAC, so all of them are included in the system. Books are a different matter since not all of the authors are IUPAC-funded scientists. There is a review process, but it is fairly easy for a book under consideration to be included with the system. If a book is being considered for inclusion, it will be approved unless there are some conclusions or views in the book that IUPAC strongly opposes. In any event, once a journal, report or book is given the approval to be included with the system, there is no need to begin a digital conversion of the material, since most of the documents were originally created in either PDF or html formats.

So now that all of the PDF and html files are created, how are they hosted? Representatives from Iupac.org went to Ibiblio.org, a not-for-profit digital archive and data repository and asked for an account with them. With help from the Ibiblio staff, IUPAC was able to set up a website that provides access to all of the journals, books and reports that they own. The setup of the chemical digital library is composed of webpages that provide listings for the digitally archived documents, with each document in the list providing a link for viewing or downloading. All of the documents are organized based on what they are (journal, book or report), and the year they were released.

Since access to the site or its materials is unrestricted, there is no need to have an account system set up for potential users. People visiting the site are only allowed to view and download documents, and they have absolutely no access to edit or remove any

documents that IUPAC hosts. Therefore, this makes administration of the site fairly simple, and the only issue that comes up every now and a again is spikes in bandwidth with IUPAC's host, Ibiblio.org, which can lead to reduced performance in the overall site. However, given the fact that the majority of material offered is in a text based format, viewing and downloading is usually fast or at relatively acceptable transfer rate, even in times of high bandwidth usage.

Iupac.org Digital Library. IUPAC's digital library, Iupac.org, offers a wide range of digital content to serve the chemistry community worldwide. Members and visitors of the site have access to daily news and announcements in the IUPAC and chemistry community, information about ingoing projects that they are funding, and summaries for sponsored conferences and symposium. However, it is perhaps the digitally archived content that is of the most interest to members and visitors alike. Currently, Iupac.org is home to 124 digital books, 265 reports, and five chemical journals with a combined total of 292 digital issues.

The reports offered at Iupac.org are available from the reports section of the site, and provide download listings that are sorted by division within the discipline of chemistry or by year. There are two hundred sixty-five total reports available from 1972 to 2004, all in PDF file format. The number of reports released each year varies from 17 to 55 reports a year. The majority of the reports are from 1996 to the present, and only 11 reports were released between 1972 and 1995 since major reporting did not start until 1996. One hundred twenty-four chemistry related books in PDF file format are also available from the site in the books section, and provide download listings that are sorted by author, year and title.

IUPAC is also home five scholarly chemical journals: *Chemistry International*, *Pure & Applied Chemistry*, *Macromolecular Symposia*, *Chemistry Educational International*, and *Solubility Data Series*, which are available from the publications section of the site. *Chemistry International* is the official newsmagazine of IUPAC which provides news about the organization, funded chemists, latest publications in the community, and upcoming conferences. *Chemistry International* is up to its 26th volume, with each volume composing a year's worth of publication, which is six issues. Currently, 42 issues in volumes 19 to 26, which are publications from 1997 to the present, are available online for download in PDF format.

The *Pure & Applied Chemistry* journal has, since 1960 strived to publish IUPAC recommendations on nomenclature, standardization, collaborative studies and data compilations. *Pure & Chemistry* is up to its 76th volume, with each volume composing a year's worth of publication, which is 12 issues. Currently, volumes 67 to 76, which are publications from 1995 to the present, have a total of 120 issues available online for download in PDF format. *Macromolecular Symposia*, another journal available at Iupac.org, publishes recent advancements and contributions in the field of macromolecular chemistry and physics. In addition they provide summaries from international meetings of IUPAC, the American Chemical Society (ACS), the European Polymer Federation (EPF), and the Society of Polymer Science, Japan (SPSJ). To date, the Macromolecular Symposia has published 214 volumes, with each volume composing of a year's worth of publication, which varies from 4 to 8 issues a year. Currently, volumes 113 to 214, which are publications from 1997 to the present, include 111 issues available online for download in PDF and html formats.

Chemistry Education International, the official newsletter of the Committee of Chemical Education, is a journal that is for high school seniors and first year college students who are seriously considering a career in chemistry. It provides aspiring chemists with facts and figures about careers in the field, as well as reports and essays from the committee regarding the current and future states of the field. Starting in 2000, the journal is up to its 5th volume, with each volume composing of a years worth of publication, which is 1 issue. Currently, volumes 1 to 5, which are publications from 2000 to the present, at total of 5 issues are available online for download in PDF and HTML formats. *Solubility Data Series*, the 5th journal offered at iupac.org, publishes compilations of all experimental determinations of solubility discovered by chemistry professionals in both the industrial and public sectors around the world. It also encourages research into projects of worldwide public interest relating to chemistry in areas such as the environment, human health, global climate change and agriculture.

V. Web Analytics

Companies and businesses invest time and money into creating and maintaining web pages in hope of generating sales or getting people interested in their products. Each organization envisions what it believes to be the best way to advertise and market their products and/or services, and the company webpage is where these ideas take form. However, the marketing of products and services does not end with the creation of the web page; in fact it only begins. Once the web page is up, organizations need to be able to somehow monitor and measure how well the website actually helps or hurts overall revenue, and this is where web analytics comes in. There are numerous metrics that can be computed to figure out how good or bad a web page is marketing products and services, and with some metrics, you can even pin-point your best and worst areas of a web page.

But, the majority of the time, these analytics are not computed by hand, they are done by specialized Information Technology (IT) firms that normally handle these requests, or a special sub-division of the company's IT, knowledge management or business management department. IT firms use web analytics packages to compute statistics. But choosing which packages to use is not such an easy decision considering the fact that there are numerous packages available on the market. Given the fact that company web pages are specially designed to meet the marketing needs of an

organization, any web analytic package that is selected must itself meet the needs of the analysts that are trying to gauge the effectiveness of the web page.

The Value of Web Analytics. Web analytics, also known as web analytics and e-analytics, is a tool that in the right hands, can be used to make cost-effective decisions. Computing analytics such as net dollar per visitor, customer drop-off rates, loyalty index, and average time spent on the system can prevent costly redesign fees associated with updating a company website. In addition, these analytics can provide significant reasons for making decisions. Currently in today's economy, requests for IT spending are usually placed under a lot of scrutiny, and without proper justification for the spending, the proposed IT project will be shelved. But with the information available from these analytics managers will have the justification they need to get the funding needed. As Susannah Patton of CIO.com stated regarding e-analytics and IT spending:

> "measuring a website's success can be crucial when CIO's are forced to defend e-business spending"(Patton 1)

But, despite the vast amount of information that is available through analyzing web traffic, many organizations do not take full advantage of what these analytics can offer. Netgenesis, a firm that specializes in web analytics, has stressed over the years how important these analytics are, suggesting they:

> "need to have as much importance as traditional accounting has in businesses…and the traffic should be as closely monitored as traditional accounting books" (Netgen 11).

Netgenesis, along with other firms and organizations that use web analytics, are a part of a growing sentiment towards the advancement of the importance of web analytics. Companies such as Amazon.com that monitor their web traffic consistently and make decisions based on those analytics tend to do better than those that do not. Keith Regan, of E-Commercetimes.com, sees the need for collecting and analyzing data to increase over the years and has stated that:

> "the web analytics sector will be worth US$1 billion annually within three years"(Regan 2).

The Best of The Best. Once a webpage is setup and is receiving regular traffic, managers will then want to purchase a package that will analyze the traffic and compute analytics, but which one should be chosen? As mentioned earlier, there are numerous packages available that can do the job, but according to E-Consultancy, a firm that specializes in web analytics and business intelligence, the general rule of thumb for purchasing a web analytics package is that large companies should purchase in-house software, and mid-size or small companies should purchase a hosted service. Based on that principle, E-Consultancy has identified what it believes are the five top leaders in providing web measurement and analytics solutions, which are the following:

1. ClickStream
2. Clicktracks
3. NedStat
4. WebAbacus
5. WebTrends

Unfortunately, the web analytics solution that Iupac.org utilizes, Urchin, is not in this list. In fact, it is not even in the top ten, according to E-Consultancy. However, it should be noted that although E-Consultancy is not the only firm that specializes in web analytics and business intelligence, they have amassed an impressive reputation over the last couple of years, and their views and opinions are valued very highly in the business intelligence community, due to their team of talented and respected analysts and scientists from the mathematical and statistics community. In any event, that is not to say that Urchin is not an exceptional application, it just means that perhaps it has not garnered enough exposure to the community to be recognized as a leader.

Clickstream is a commercial log analysis tool that monitors a visitor's clickpaths, which is the route that visitors choose when navigating or "clicking" through a site. At any given time, a web analyst can see all the pages viewed by any given visitor, presented in their succession of mouse clicks, which is the order the pages were viewed. From this information, an analyst can tell when and where a person came in to a site, all the pages they viewed, the time they spent on each page, and when and where they left. Like all commercial log analysis tools, they are on average more accurate and supply significantly more information, but at a high price.

Clicktracks is another commercial log analysis tool that monitors clickpaths, but it has two distinctive features that Clickstream does not: it is highly customer friendly, creating graphs and tables from log data & being very graphically oriented, and it costs much less and in some respects does a lot more. The package that was designed because

of frustration with existing website analysis tools, and as a result appears to be a great success not only in the US, but in Europe as well.

Nedstat is a subscription-based, hosted web analytic tool. This means that an online server runs the package and all of the information it records is kept on a server. Because the service is hosted, the package can provide you with information 24 hours a day in real-time. Nedstat essentially provides the same level of service as Clickstream, however since it is a subscription based service, the price is significantly cheaper in the short run than any commercial log analysis tool.

WebAbacus is another commercial log analysis tool that monitors user clickpaths. It provides the same features as Clickstream and Clicktracks, but it does one thing that the others do not; it provides possible routes of action in e-commerce, marketing, and knowledge management based on what information is tracked in the log files. This feature makes this package live up to its name.

WebTrends is again another commercial log analysis tool that monitors and logs user activity on WebPages. Somewhat similar to WebAbacus, it provides information in detailed segments such as e-commerce, marketing, merchandizing, content and navigation, and visitor segmentation, but it does not provide advice or possible courses of action based on what user activity has been recorded. In addition, it is also the only one of the five analytics tools that can use web tags that are placed inside of pages to track users, instead of the usual cookie tracking systems.

Choosing the best web analytic package will make it much easier to determine what should be done to Iupac.org in order to make the site better for its intended audience of chemistry enthusiasts and professionals.

Findings Based On Comparison. After researching each package by going to its website and reading reviews on its performance, it is easy to see that most packages offered the same kinds of services, and there was always one service that they excelled in or put the most emphasis or effort towards.

Nedstat was the only service that was an Application Service Provider (ASP) only hosted package, meaning that the software does not need to be installed on a machine, but this also meant that if the server hosting the service was down, the service would not be logging all activity, and that is what was frequently happening, according to reviews of customers. Although Nedstat performs most of the standard features of web analytic tools such as reporting, analyzing and monitoring web traffic, and its two areas of excellence are low price and multi-language support. But there are also issues about whether or not the data being displayed is correct because of lapses in web monitoring due to server downtime.

ClickStream and WebTrends both have the best tracking abilities. ClickStream has the ability to track user activity online or offline, making it theoretically possible to track everything that happens on the webpage. Although WebTrends cannot track user activity offline, it is more privacy-friendly towards tracking user activity, and is the only one that can track activity using tags inserted into WebPages to gather data from visiting packages and collect the information in a specified central location. This method of

recording user activity is much more efficient than gathering web logs from various servers and other locations. Which is better for a website depends on the preferences for information retrieval and system architecture. Using tags in WebPages to monitor user activity is certainly a more efficient way of monitoring, but if the architecture does not support this, then Webtrends has lost its advantage.

ClickTracks, was the leader in path analysis, which is strengthened by their emphasis on simplicity and straightforwardness. This is done by making as many screens and features utilize as much graphical content as possible. Their main goal in designing a system was to make it so that anyone could understand it, since there have been considerable complaints about these systems being too complex and hard to understand. The last package, WebAbacus, features robust reporting power, which is its biggest strength. It has a reputation of being able to provide powerful custom reporting and data manipulation, automatic alerts for analytics that need to be addressed based on monitoring, and the ability to provide suggestions for courses of action based on recorded data.

Now that all of the top five clients have been introduced and compared which is the best web analytic tool available? Any of the four web analytic tools, ClickStream, ClickTracks, WebAbacus, and WebTrends are worthy solutions, so the choice depends on what the focus for the your system. If user tracking is most important, then ClickStream or WebTrends are better. If powerful reporting abilities are most important, then WebAbacus is better. And if the ability to be able to thoroughly analyze webpage navigation paths then ClickTracks offers the better solution.

Challenges Associated With Analytics. Although the potential benefits of utilizing web analysis software appears to be limitless, further research into the current state of the field proves otherwise. There are four challenges that are associated with web analytics. The first is an overall lack of analytical professionals to analyze the results. Just because the software is generating results does not necessarily mean that anyone can interpret the results, despite what the vendors say. An individual who has the ability to consider all the possibilities behind why certain results were computed is needed, since the same set of data can be interpreted hundreds of different ways. It is impractical to purchase any one of these packages without someone who is competent in that area to properly analyze the data.

The next challenge is industry standards for web analytics. Currently, there are no benchmarks for excellence in the field, so when organizations analyze the results generated by their web analysis applications, they have nothing to compare or test it against. This can prove to be frustrating considering the amount of time and money invested into computing web metrics. Although there are annual conferences and summits on what constitutes excellence in the field, scientists and vendors from all over the world have not been able to agree on a set of standards. And not having an industry wide standard for excellence makes upper management wary of investing in these applications.

Integration of the massive amounts of data generated is another challenge. In most cases, data will have to be integrated from various sources to accurately analyze web site activity. And finally, the number one challenge is trying to advertise the value of these applications to upper management. Web metrics software vendors have to sell

organizations on the idea of how these applications can not only benefit them, but also generate a profit at the same time. And given the three previous challenges just mentioned, it can be quite difficult. Typically management is willing to take calculated risks as long as the benefits are worth it, but given the fact that the value of these applications can vary from organization to organization, it might be a risk that management is not willing to take.

VI. Urchin Web Analytics Software

Urchin v.5, the web analytics package that is used by Ibiblio.org, which is the host of Iupac.org, retails for $895.00, which includes the base license for use of the package. It is primarily intended for Internet Service Providers (ISP) and large corporate web sites to handle all log analysis for generating web statistics. It features a browser based online reporting system, multilingual functionality, and runs on 15 different variants of UNIX, Windows NT, Windows 2000 and even Cobalt, a legacy programming language.

Urchin, despite its lack of notoriety, provides features and levels of functionality that are beyond the details of most other web analytics packages. Because of this, the information provided by Urchin gives management and clients alike the ability to make informed decisions about how to advertise or organize content on their web sites. Most web analytic software packages are able to provide detailed reports on web statistics areas such as the total number of page views, hits, visitors, bytes transferred, entrance pages visited, exit pages visited, top pages visited, search engine keywords used, browser types, computer types and referrals.

Urchin provides all of those features and takes it a step further with their first party cookies, called Urchin Tracking Modules (UTM). UTMs track all kinds of data from the visitors of a site, and provide detailed information on user activity that is beyond the levels of data that are normally logged with simple web server log files. For example, in Urchin, actions of users are categorized in tasks and goals, where a specific set of tasks

encompasses a goal. Goals in Urchin can be tracked rather easily, such as number of times a visitor comes to a specific page and the length of time it takes each user to reach that particular goal. In addition, with the E-Commerce and the Campaign Tracking Modules, which are part of the UTM, each web page on a site can be fine tuned to increase the overall success rate for that individual page and the entire site.

How a visitor stumbles upon a site is just as important as what actions the visitor took after arriving at the site. Referrals, which are the methods to which a visitor discovers a site, can prove to be very useful in determining user actions in Urchin. The information that Urchin processes from each referral can provide invaluable data on where the user came from, such as another website or a search engine. Additionally, the information provided from each referral can be further processed to determine which web sites and search engines brought a site the most traffic and even which keywords the visitor used on a search engine to find a site.

In the end, most web sites need analytic software packages like Urchin that can tell them how and why visitors are coming to their sites, and what they are doing once they get there. And knowing how and why visitors behave in specific ways is a benefit that is invaluable to an institution or organization.

Benefits. General benefits of using any web analytics package are the following:

- Automated report generation, which means no need to manually produce reports;
- Sorting and filtering of reports, making it easier to search for information about visitors;

- User and referral tracking, which helps determine what users are doing on a website, and where they are coming in from;
- Search engine reports, which helps determine which search engines provide traffic to a website and what are the keywords that visitors used while trying to search for a website; and
- Real-time web statistics, to see what is happening on website as it happens, as opposed to hours or sometimes days later.

The specific benefits of using Urchin however mainly affect management, web developers and IT staff responsible for maintenance on web servers.

For management, the first benefit is reducing the overall management burden, because Urchin in a sense coordinates integration of all of the different data tracking processes to produce accurate reports. Imagine for example each data tracking process is a department or team in a company. After tens, maybe even hundreds of these departments finished collecting data, they would then be going to a department head or manager to coordinate the integration process to create one massive report. That is no small feat, and the management overhead for an integration of that size would be massive. But with Urchin it is a simple task.

The ability to accurately see short-term and long term trends is the next benefit for management. Urchin has the ability to provide numerous web metrics over specific ranges of time selected by a user. This allows management to see exactly how that new search engine that was implemented 6 months ago has done in the first month is was in service. Being able to look at initiatives over the short-term and long term is something

that is highly valued by management, and the built in features in Urchin make it as easy as pointing and clicking to find this information. And finally, learning more about an organization's primary visitors and how to retain them is the last benefit to management. The detailed reports provided by Urchin allow management to see who is normally visiting the website, and specifically where and when visitors are coming in and leaving. Information like this makes it much easier to potentially focus in on areas of the site that generate the most traffic and referrals, and areas that are generating the least amount, to maximize traffic on all levels of the site.

Web developers are the next group of people that Urchin specifically benefits, and the first benefit for them is demonstrating the value of improvements to usability of the site. Say for example a developer for the company website feels that if they add and/or delete certain elements of the structure of the site, overall usage of the site will drastically increase. These changes can be implemented and then after a month or two, management can look at the usage levels before the changes were implemented and after, and then make a decision as to whether or not the changes suggested need to remain or be removed. It is really a remarkable tool, because nothing is more of a factor in the decision making process than indisputable facts, which is something that Urchin can easily provide. The other benefit to web developers is how design decisions can improve a website's overall return on investment. Increases in usage are great for a site, but organizations that conduct a lot of business through e-commerce, will want to see which people visiting the site are purchasing goods or services. Using the E-Commerce modules mentioned earlier, Urchin makes it very easy to do this, and the information

provided in usage and e-commerce reports will make it even easier to decide how good or bad a decision was.

IT staff responsible for web server maintenance is the last group of people that directly benefit from Urchin. The first and maybe the most useful benefit is being able to track down and fix outdated or broken links, redirect errors and referral errors. Again, given the fact that Urchin is an all purpose tool, IT staff can use one single tool to do all of these tasks, as opposed to either using free or fee-based online or offline tools to do this task. It is a nice bonus to the people who are considering which web analytics software package to buy, since it provides so many useful web site maintenance tools in addition to the main statistical reporting features which are standard. Minimizing server resources for web analytics is the next benefit to IT staff. Again, with Urchin being a single all-in-one solution to a statistics problem, the IT staff will not have to run numerous statistical applications to get the same detailed information. And this in turn means an overall reduction of server resource overhead, which is good for the IT people and good for management, since they will not have to buy any more servers to meet the demands of the statistics initiative.

The final benefit for IT staff is the ability to monitor server throughput and bandwidth usage. Again, Urchin provides many useful tools outside of the standard statistical ones, and utilities for monitoring bandwidth and site usage make it even more attractive to institutions and corporations alike because it can benefit so many people in an organization, not just a small group of people. As IT maintenance people, it is assumed that they would not entirely rely on the bandwidth and server utilities that are provided with Urchin, but they would be grateful that they are available nonetheless, and

could be possibly used as something to compare their own results from other bandwidth and usage utilities that they use.

Features. Urchin Software Corporation has provided a very detailed and exhaustive list of all features and functions at http://www.Urchin.com/products/v5/feature_list_master.html. A complete list of the features of Urchin is found in appendix a.

Comparison Against Competition. So, how does Urchin, which is marketed as an enterprise level application at a mid-market price, stack up against other web analytics packages? Well, to answer this question, the Urchin Software Corporation first came up with 13 aspects that they feel a successful analytics package should have, which are the following:

1. 1st Party Cookies
2. Software & On Demand
3. Wide Platform Support
4. Datacenter-class Scalability
5. Web Based Reporting
6. Site Overlay without Plugin
7. Patents Issued & Pending
8. Auto-Import of CPC Cost Data
9. Click Fraud Analysis
10. Paid vs Organic SEM Analysis
11. Dynamic Visitor Segmentation

12. Visitor & Content Scoring

13. Cross Channel Integraton

Then, in order to make this a fair comparison, they picked other packages that they felt were in the same league as them, which are the following: WSS HBX, NetIQ WebTrends, Core Metrics, ClickTracks, and Omniture SiteCatalyst. Ironically, 2 out of the 6 chosen, WebTrends and ClickTracks, are in E-Consultancy's top 5 web analytics packages, which was mentioned earlier. Finally, they provided a competition matrix which shows how they compare against these 6 packages, which is displayed below:

	Urchin	WSS HBX	NetIQ WebTrends	Core Metrics	Click Tracks	Omniture SiteCatalyst
1st Party Cookies	Yes	No	Yes	No	No	No
Software & On Demand	Yes	No	Yes	No	No	No
Wide Platform Support	Yes	n/a	No	n/a	No	n/a
Datacenter-class Scalability	Yes	No	No	No	No	No
100% Web Based Reporting	Yes	No	No	Yes	No	Yes
Site Overlay without Plugin	Yes	No	No	No	n/a	Yes
Patents Issued & Pending	Yes	Yes	Yes	No	No	No
Auto-Import of CPC Cost Data	Yes	No	No	No	Yes	No
Click Fraud Analysis	Yes	No	No	No	No	No
Paid vs Organic SEM Analysis	Yes	No	Yes	No	Yes	Yes
Dynamic Visitor Segmentation	Yes	No	Yes	No	No	No
Visitor & Content Scoring	Yes	No	No	Yes	No	Yes
Cross Channel Integraton	Yes	Yes	Yes	Yes	No	Yes

Fig. 4. Urchin Competition Matrix, Urchin.com 25 October 2004: http://www.urchin.com/products/v5/matrix_competition.html

Although it is somewhat of an ironic coincidence that all of the 13 aspects that Urchin feels is essential are features that it has standard, they are indeed important things

that should be addressed. Looking at the results, it appears as if WebTrends is their strongest competitor. But, regardless of the competition results and the criteria selected, based on what the package has to offer, Urchin is a force to reckoned with in the industry, and despite not being on everyone's top 5 or even top 10 list, it is an exceptional web metrics solution.

VII. Web Survey with phpESP

The best way to approach answering the question of what should be done to iupac.org in order to make the site better for its intended audience of chemistry enthusiasts and professionals would be to do a survey of users to gauge the overall satisfaction level. After a bit of searching, phpESP, an open source survey tool emerged as being the best option to create and distribute the survey.

phpESP provides a way to create and administer web-based surveys online. It also has an administrative component which allows owners of surveys to view the results of each survey. In addition, every time a respondent submits a completed survey, an email notifies the owner of that survey. As stated earlier, phpESP anonymously stores participant responses in an online database. Each response is identifiable only by a unique survey number, which can in no way identify the specific respondent who completed the survey.

Once the survey was completed, the original 25 people who were used in a usability test for a search engine on Iupac.org were contacted again via email and asked to participate in the survey. The survey officially went online on October 7, 2004, and stayed up for a month, going offline on November 8, 2004. Out of the 25 people

contacted, 17 responded back.

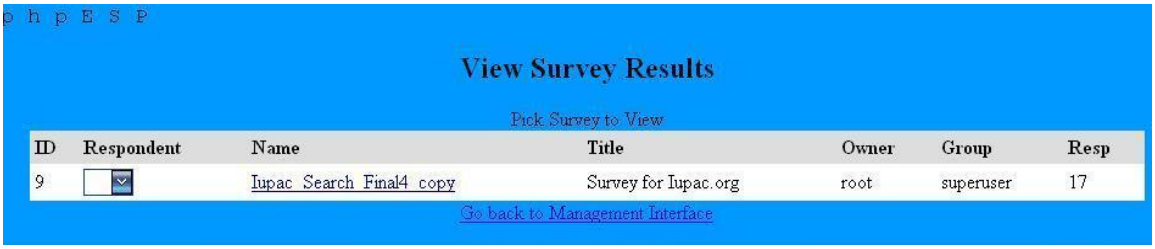


Fig. 5. phpESP survey results.

Questions. When the survey went online on October 7th, the following questions were included:

1) Are you a member at IUPAC?

1 - Yes

2 - No

2) How often do you visit Iupac.org?

1 - rarely

2 - sometimes

3 - often

3) How often do you use the search function at iupac.org?

1-never

2-rarely

3-sometimes

4-often

4) How would you rate the search function?

1-poor

2-fair

3-good

4-excellent

5) Do you have any problems using the search function?

-open ended

6) Do you feel that the search function returns the appropriate results?

-open ended

7) How many times do you usually have to do a search before you find what you are looking for?

1- 1 to 2 times

2- 2 to 5 times

3- 5 to 10 times

4- more than 10 times

8) How many pages of results do you usually go through before you find what you are looking for?

1- 1 to 2 pages

2- 2 to 5 pages

3- 5 to 10 pages

4- more than 10 pages

9) When you run searches and the results are not what you wanted or expected, what do you think is the cause?

1- the terms used in my search were not sufficient

2- the search function itself

3-other

10) Are there any things you would like improved with the search function?

-open ended

11) How would you rate the content available at IUPAC?

1-poor

2-fair

3-good

4-very good

5-excellent

12) What sections of Iupac.org do you visit the most?

-open ended

13) Do you think that it is easy to navigate through the site to find what you're looking for?

1- yes

2- no

14) Are there any sections of the site that could be improved?

-open ended

15) Any other suggestions or comments about iupac.org?

-open ended

Here are the results of each question, followed by initial comments and observations on the results.

1. Are you a member at IUPAC?		
Yes	58.8%	(10)
No	41.2%	(7)
TOTAL	**100%**	**17**

Table 1. Question 1 results.

As predicted, more members of IUPAC participated in the survey, but it is surprising that the percentage between members and non-members would be so close, as it is 58.8% to 41.2%.

2. How often do you visit Iupac.org?		
Rarely	5.9%	(1)
Sometimes	41.2%	(7)
Often	52.9%	(9)
TOTAL	**100%**	**17**

Table 2. Question 2 results.

The majority of participants, 52.9% visited the site often, while 41.2% sometimes visited the site. One respondent said that he visits Iupac.org rarely was a member of IUPAC. This was determined after cross tabulating the overall results of the survey and selecting to view how members responded and how non-members responded. Cross tabulating the results based on member affiliation, 70% of members visited the site often, while 71.4% of non-members sometimes visited the site.

3. How often do you use the search function at lupac.org?		
Never	5.9%	(1)
Rarely	29.4%	(5)
Sometimes	52.9%	(9)
Often	11.8%	(2)
TOTAL	**100%**	**17**

Table 3. Question 3 results.

The majority of the participants, 52.9% sometimes used the search function, while 29.4% rarely used the search function, and 11.8% used the search function often. The one respondent who said that he never uses the search function was a non-member. This is not much of a surprise, but then again, one would assume that since the respondent that said that they rarely visit the site was a member, it can also be assumed that the respondent that said that they never use the search function would also be a member, if not the same respondent who answered "rarely" for question 2. Cross tabulating the results based on member affiliation, the majority of members, 50.0% sometimes used the search function, and the majority of non-members, 57.1 also sometimes used the search function.

4. How would you rate the search function?		
Poor	5.9%	(1)
Fair	11.8%	(2)
Good	82.4%	(14)
Excellent	0.0%	(0)
TOTAL	**100%**	**17**

Table 4. Question 4 results.

As predicted, the majority of respondents thought the search engine

was good. But it is interesting that no one thought that it was excellent. The percentages for fair and poor ratings for the search function are not a surprise. In any event, after cross tabulating the results based on member affiliation, it was found that the majority of members felt that the search engine was good, whereas all of the non-members felt that it was good as well.

5. Do you have any problems using the search function?	
#	Response
1	Disentangling publications hits from the rest
1	don't use it
1	don't use it that much
1	e.g. Stibr (PAC 2003, 75(9), 1239-1248)
1	having to remember to select the right option from the menu
1	I forget to select the correct area to search for when searching
1	I hardly use it, but I guess no
8	no
1	often want to search nomenclature & terminology, which does not work
1	Search should be highlighted by a search box from every screen

Table 5. Question 5 results.

Question five is the first of the open ended questions. From the results, it would seem as if the results are split down the middle 50/50 for those having problems with the search problem and those that are not. Cross tabulation of results by member affiliation showed that members were also split 50/50 on the subject of problems with the search function. Non-members were slightly different, with the majority of respondents stating that they did not have problems with the search function.

6. Do you feel the search function returns the appropriate results?	
#	Response
1	don't know
1	Generally yes
1	I don't know
1	I guess
1	most of the time
1	not all the time, but mostly
1	often want to search nomenclature & terminology, which does not work
1	Typically too much information - too many hits
7	yes
1	Yes but would be better if searching could be more selective
1	yes I think so

Table 6. Question 6 results.

As expected, it appears as if the majority of respondents to question six felt that the search function does indeed return the correct results. Cross tabulation of results by member affiliation showed that 77.8% percent of members felt that the search engine does yield correct results. Out of the non-members, after throwing out 2 responses("don't" and "I don't know"), 5 of them felt as if the search did indeed return the correct results.

7. How many times do you usually have to do a search before you find what you're looking for?		
1 to 2 times	35.3%	(6)
2 to 5 times	58.8%	(10)
5 to 10 times	0.0%	(0)
more than 10 times	5.9%	(1)
TOTAL	**100%**	**17**

Table 7. Question 7 results.

Results here were nothing out of the ordinary, as the majority of respondents ("1 to 2 times" + "2 to 5 times") found what they were looking for after performing 1 to 5

searches. Cross tabulation of results based on member affiliation showed that the majority of members, 60%, performed a search 2 to 5 times, and the same applied to the non-members.

8. How many pages of results do you go through before you find what you're looking for?		
1 to 2 pages	41.2%	(7)
2 to 5 pages	41.2%	(7)
5 to 10 pages	11.8%	(2)
more than 10 pages	5.9%	(1)
TOTAL	**100%**	**17**

Table 8. Question 8 results.

The majority of respondents searched through 1 to 5 pages of results to find what they were looking for. It is interesting to note that the numbers were split so evenly for the majority response between "1 to 2 pages" and "2 to 5 pages". Cross tabulation of results based on member affiliation showed that the majority of members searched through "2 to 5 pages" of results, and the same applied to the non-members

9. When you run searches and the results are not what you expected, what was the cause?		
The terms used in my search we not sufficient	41.2%	(7)
The search function itself	41.2%	(7)
Other: too many entries in archives have same key words	11.8%	(2)
TOTAL	**100%**	**17**

Table 9. Question 9 results.

Question nine yielded some very interesting results. It was predicted that the majority of respondents would blame the search itself for not returning the correct results, however it appears as if the results are almost straight down the middle, with 5.9% of

respondents neither blaming themselves nor the search function itself. And after cross tabulation of results based on affiliation, it was discovered that the majority of members felt that it was the search terms used that were the main cause for undesired results, whereas the majority of non-members felt that it was the search engine itself that was at fault for returning undesired results. The one person who blamed undesirable results on the content that is hosted on Iupac.org was a member.

10. Are there any things you would like improved with the search function?	
#	Response
1	a different way to do searches
1	don't use it
1	enable nomenclature & terminology to be searched
1	I would prefer not to change automatically the search area when search is done
1	instructions on how to use
8	no
1	not at the moment
1	Not sufficiently skilled to make recommendations
1	search within results
1	See answers 5 and 6

Table 10. Question 10 results.

After throwing out 3 responses that did not seem to answer the question ("See answers to 5 and 6", "Not sufficiently skilled to make recommendations", and "don't use it), the majority of respondents felt that there was nothing that needed to be improved with the search function. Cross tabulation of results based on member affiliation showed that members were split 50/50 on the subject of whether or not things needed to be improved with the search function, while the majority of non-members felt that no improvements needed to be made to the survey. However, all suggestions will be taken into mind when looking at recommendations for improving the search engine function.

11. How would you rate the content available at IUPAC?		
Poor	0.0%	(0)
Fair	5.9%	(1)
Good	47.1%	(8)
Very good	17.6%	(3)
Excellent	29.4%	(5)
TOTAL	**100%**	**17**

Table 11. Question 11 results.

The majority of respondents had favorable views of the content provided by IUPAC. Cross tabulation of results based on member affiliation showed that the majority of members felt that the content at Iupac.org was good, whereas the majority of non-members felt the content was either very good or excellent. It is an educated guess that the members of IUPAC are harder to please as far as content is concerned, and that may be why they were less impressed with the content than non-members.

12. What sections of Iupac.org do you visit the most?	
#	**Response**
1	Analytical Chemistry Division
1	c.e.i.
2	CEI
2	CEI publications
1	CPEP, projects
1	many
1	membership lists
2	PAC
7	project
1	Projects
1	readers corner
1	the reports section

Table 12. Question 12 results.

The specific area that respondents seemed to visit the most, was the CEI publications section. However, considering the fact that CEI publications is under the

Readers Corner area, as well as Reports and PAC, one could also say that the most visited section of the site is the Readers Corner area. Cross tabulation of results based on member affiliation showed that the members visited the Reports section the most, while the majority of non-members visited the CEI area of the site.

13. How easy is it to navigate through the site to find what you're looking for?		
Very easy	0.0%	(0)
Easy	5.9%	(1)
Not easy but not difficult - somewhere between	47.1%	(8)
Moderately difficult	17.6%	(3)
Difficult	29.4%	(5)
TOTAL	**100%**	**17**

Table 13. Question 13 results.

As expected most respondents, answered that the navigation of the site was not quite easy or difficult, but rather somewhere in between. Surprisingly cross tabulation of results based on member affiliation showed that the majority of non-members felt that site navigation was easy, whereas most members felt that the navigation was neither easy nor difficult, but rather somewhere between.

14. Are there any sections of the site that could be improved?	
#	Response
1	-
1	All of them!
1	I am not expert enough to offer a worthwhile opinion
1	I can think about it!
1	It is being regularly updated but there is a dependency on task groups
4	no
1	organization is poor
1	organization of all publications. It took me awhile to find everything
1	search engine for nomenclature & technology
1	The navigation menu at the main page. It needs to be more descriptive
2	The publications section. Show all publications offered from main page
1	Would prefer PAC available as HTML
1	you should display volume indexes in a better format

Table 14. Question 14 results.

After looking at the results and throwing out some answers that did not answer the question ("I am not expert enough to offer a worthwhile opinion" and "I can think about it!"), it appears as if most respondents did not think any sections of the site could be improved. Cross tabulation of results based on member affiliation showed that members felt that the overall organization of the site needed to be improved, and non-members felt that the publications section needed to be improved. However, all suggestions will be taken into mind when looking at recommendations for improving the site.

15. Any other suggestions or comments about Iupac.org?	
#	Response
1	-
1	a great place
1	a great place for chemistry materials
1	good job iupac
1	just the navigation
1	needs links to; databases; downloadable software; access to order form
6	no
2	None
1	S. this is Fabienne testing the survey ;->
1	search engine for nomenclature & terminolgy
1	The site\'s design is awful and should be modernized

Table 15. Question 15 results.

Given the fact that this last question was a general suggestions or comments question, the varying results are not a surprise. Participants suggested or commented about the navigation, the search engine, and the overall site design. Again, as mentioned earlier, all suggestions and comments will be taken into mind when looking at recommendations for improving the site.

Preliminary Observations. So, what does this survey say about the users of Iupac.org? Most visitors of the site are indeed IUPAC members, and they use the search engine function somewhat on a regular basis, either using the function sometimes or often. Most visitors, both members and non members alike have favorable views of the search engine, but still have issues with it. Most also felt that the search engine returns correct results, and can generally find what they are looking for after performing 1 to 5 searches and searching through 1 to 5 pages of results. Interestingly, members and non-members were split on the issue of who is at fault when the search engine returns undesired results, since the majority of members felt that the keywords used for the search was responsible for unwanted results, while the majority of non-

members felt it was the search engine itself. Ironically, the majority of respondents did not feel that the search engine needed to be improved, but there were some who suggested that instructions on how to use it are needed.

As expected, most respondents had very favorable reviews of the content available at Iupac.org, although non-members thought more favorably of the content than the majority of members. As also expected, since Iupac.org is home to so many chemistry related publications and books, it was no surprise that a vast majority of respondents visits the Readers Corner section the most. Responses about navigation were also as expected, with more respondents either finding it easy or somewhere in between easy and difficult to navigate through the site to find what they are looking for. As far as improvements to the site, the suggestions did vary, with people mostly suggesting that the organization of the site and the format of specific publications be improved.

VIII. Iupac.org Web Statistics

Now that the voice of the users of Iupac.org, both member and non-member was heard, it is useful to see if what they said about their experiences on the site was the same as what they were doing on the site. Once the survey results were analyzed and compiled, the next step was to analyze the web server logs to determine user behavior on the Iupac.org. User activity was limited to October 7, 2004 to November 6, 2004, roughly around the time when the web survey was distributed to the participants of the study. Before beginning, here are a few terms that one should know before looking at these statistics:

- **Session**: A series of hits to the site over a specific period of time by one visitor.
- **Pageview**: A request to the web server by a visitor's browser for any web page; this excludes images, JavaScript, and other generally embedded file types.
- **Hit**: Any successful request to a web server from a visitor's browser.
- **Bytes**: The quantity of network bandwidth used by the files requested during a specific date range.

The first area that was investigated was where were the visitors coming from, specifically top-level domain wise. These top level domains are names that identify a group of Internet Protocol (IP) addresses.

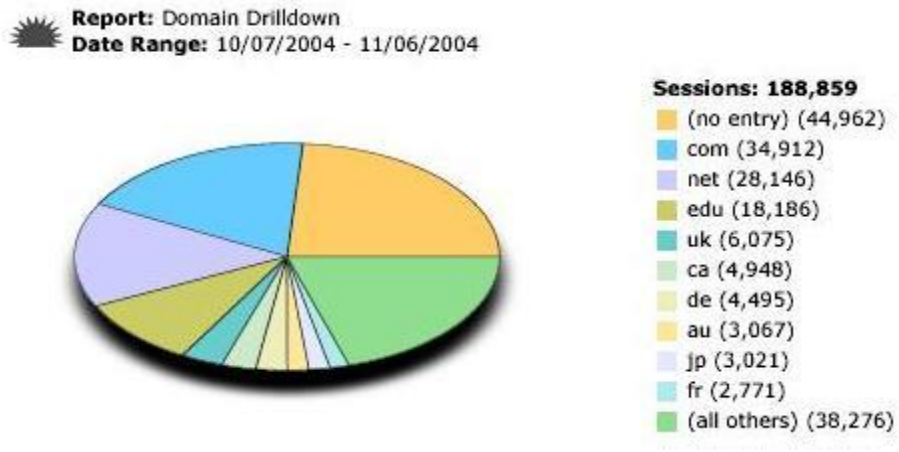


Fig. 6. Internet domains.

Domains. Looking at the results (figure 6), excluding visitors who had domains that were not able to be identified ("no entry" on the chart), most of the visitors were from the .com, or commercial business domain name, with .net (network organizations) and .edu (educational institutions) closely following. However, it is no surprise that since IUPAC is a world renowned organization that the rest of the top level domains are all European or Asian domains, such as the United Kingdom, Germany, Australia, Japan and France. Moving down from the top level domains to the lower level domains of organizations or companies that host internet service shows the following (figure 7):

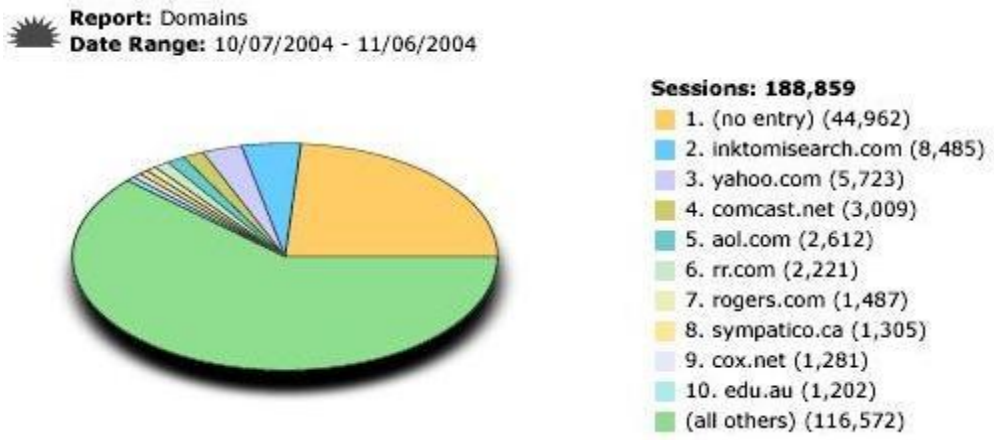


Fig. 7. Lower level internet domains.

Excluding the “no entry” results for users who had domains that were unidentifiable, it would appear as if users from the “linktomisearch.com” were the most frequent visitors of the site, with people from Yahoo.com and Comcast.net following. It is interesting to note that only two out of the top ten domains are foreign domains, which is somewhat of a surprise, given the fact that IUPAC is a worldwide organization. However, it should be noted that this does not mean that all of the foreign visitors of the site only came from those two domains. It is very possible that any one of the .com or .net domains could be providing internet service to not just domestic visitors, but foreign ones as well.

Countries (1-10) / 149	Sessions	Percent	0 23.81%
1. (no entry)	44,962	23.81%	
2. com (Commercial)	34,912	18.49%	
3. net (Network)	28,146	14.90%	
4. edu (Educational)	18,186	9.63%	
5. uk (United Kingdom)	6,075	3.22%	
6. ca (Canada)	4,948	2.62%	
7. de (Germany)	4,495	2.38%	
8. au (Australia)	3,067	1.62%	
9. jp (Japan)	3,021	1.60%	
10. fr (France)	2,771	1.47%	

Fig. 8. Countries visiting Iupac.org.

Looking at the top ten countries of visitors that come to Iupac.org (figure 8), It would appear as if The United Kingdom, Canada, Germany, Australia, Japan and France are home to IUPAC's most frequent users. But then again, it is possible that a large percentage of the .com, .net and .edu domains could be U.S. visitors.

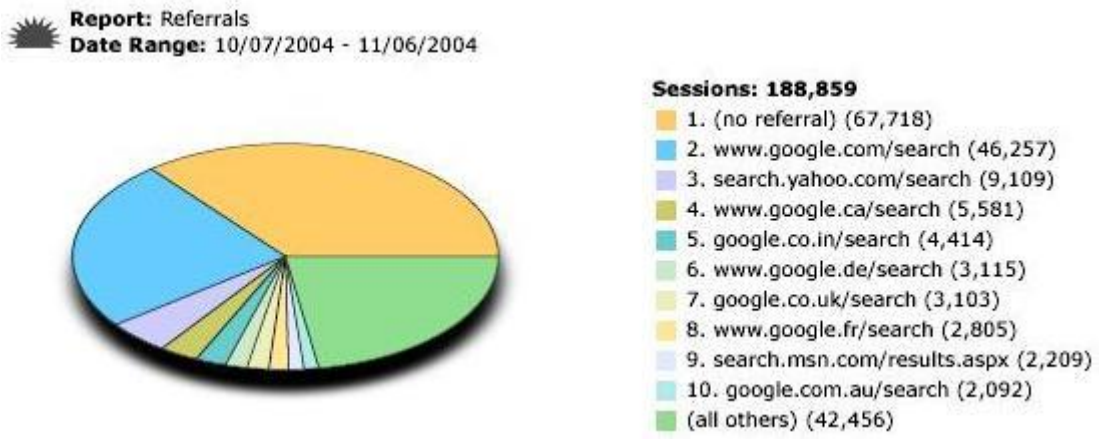


Fig. 9. Referrals.

Referrals. Referrals are URLs that bring traffic to a website, whether it be a website or a search engine. Looking at the results, it would appear as if the majority of referrals are Google searches from the U.S., Canada, United Kingdom, France and Australia, with the rest of the referrals coming from Yahoo and MSN searches.

Search Terms (1-10) / 39,493	Sessions	Percent	0 29.85%
1. (other)	31,813	29.85%	
2. iupac	1,676	1.57%	
3. rf+value	355	0.33%	
4. pure+and+applied+chemistry	255	0.24%	
5. ionic+strength	253	0.24%	
6. relative+standard+deviation	218	0.20%	
7. isosbestic+point	172	0.16%	
8. collision+theory	169	0.16%	
9. amphipathic	153	0.14%	
10. graphene	149	0.14%	

Fig. 10. Search terms used in referrals.

Not surprisingly, the top ten search terms used in referrals, shown in figure 10, shows that the number one search term (excluding the “other” result at number one) is “iupac” and that “pure and applied chemistry” is number three. Naturally this makes sense, and it should be noted that since the majority of respondents stated that they visited the “Readers Corner” section of the Iupac.org the most, which is composed of publications, with “pure and applied chemistry” or PAC being one of them, it makes sense to see “pure and applied chemistry as one of the top ten search terms. In addition, after looking through the first 100 search terms, it would seem as if users use between two and three words in queries, and rarely use any boolean conditionals like “or”, “and”, or “not”. However, “other” did have the highest percentage in search terms used. Terms

that are identified as other are terms that are not strings, so it is possible that the "other" searches could be searches where a visitor did not type in a search term.

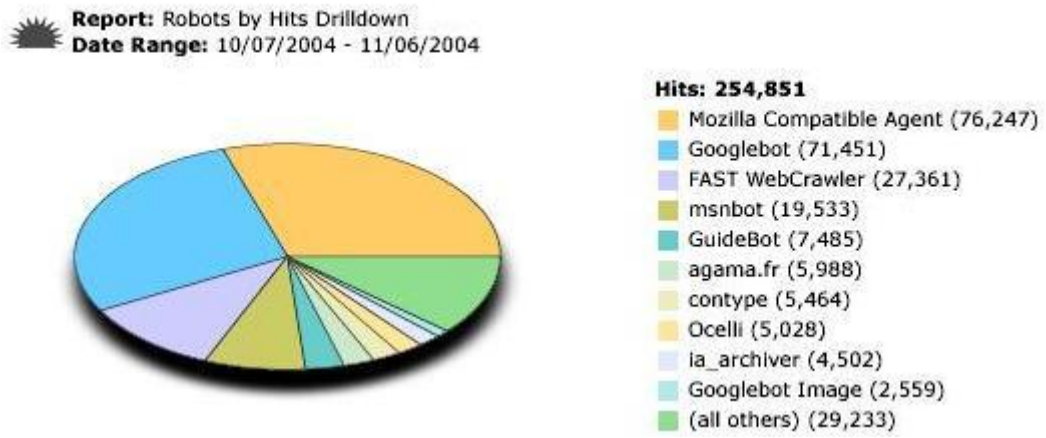


Fig. 11. Robots.

Robots. Robots are the spiders or "web-crawlers" that are used to search content in a domain and return the results to a search engine for reference. Search engines frequently use this technology to determine what a website has to offer, and then indexes the data from the web-crawling to use with their searches. Figure 22 shows the top ten robot technologies used on Iupac.org. Not surprisingly, the Google robot technology, Googlebot came in at number two, while the Mozilla Compatible agent came in first. Seeing MSNbot at number four is not a surprise either, considering the fact that MSN search was number nine in top ten referrals to Iupac.

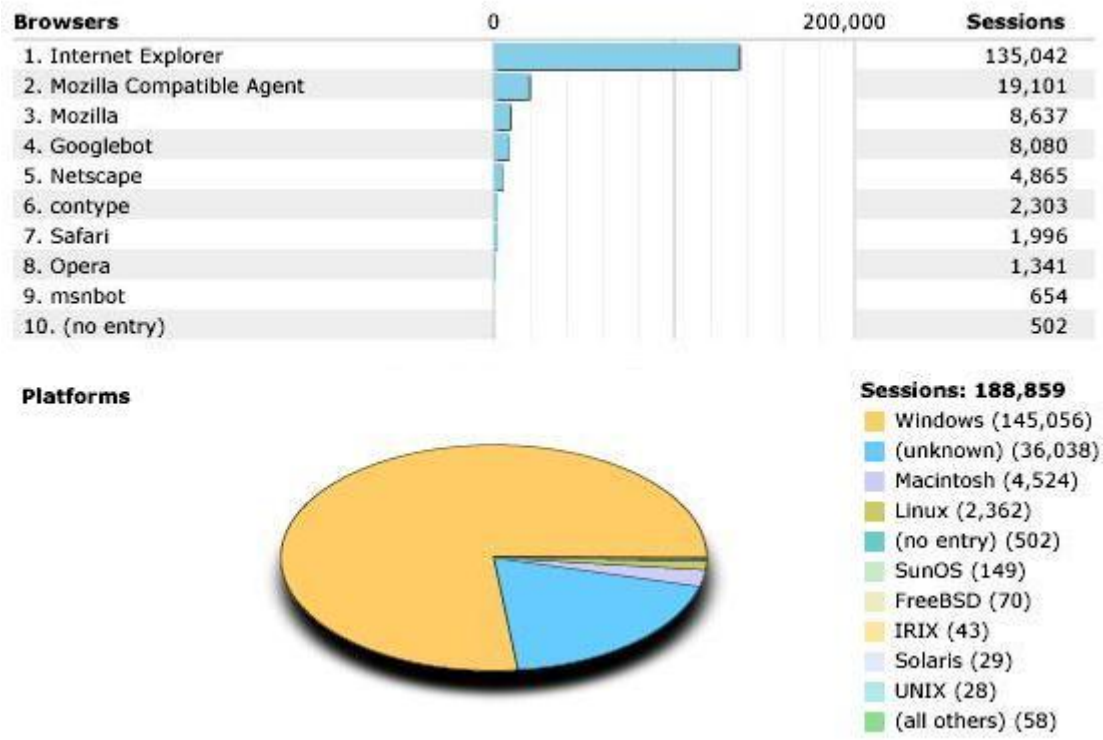


Fig. 12. Browers and platforms.

Browsers. Users of Iupac.org primarily used Internet Explorer as their web browser, with Mozilla and Googlebot coming in at second and third. Given the popularity and success of Microsoft's Internet Explorer, this should be expected. The majority of platform used by users was Windows, another Microsoft product, with Macintosh at a distant third. However, it is interesting to find the number two platform was "unknown". More than likely, if an operating system is not Windows or Macintosh, it is some variant of UNIX, like Linux. Seeing how the rest of the platforms listed were either UNIX or some flavor of Linux, it is probable that the "unknown" platform was probably a variant of Linux that was unable to be detected.

Fig. 13. Sessions.

Traffic. The first aspect of traffic that was looked at was the number of sessions that normally occur during the day. As mentioned before, a session is a series of hits to your site over a specific period of time by one visitor. Based on the results in figure 24, the average number of sessions per day between October 7, 2004 and November 7, 2004 is 6,092.23, with the lowest sessions per day being 2,621 and the highest being 9,271. One thing to note here is the pattern of total sessions during the week. Based on the results, it appears as if more people visit the site during the middle of the week, and less people visit it during the weekends.

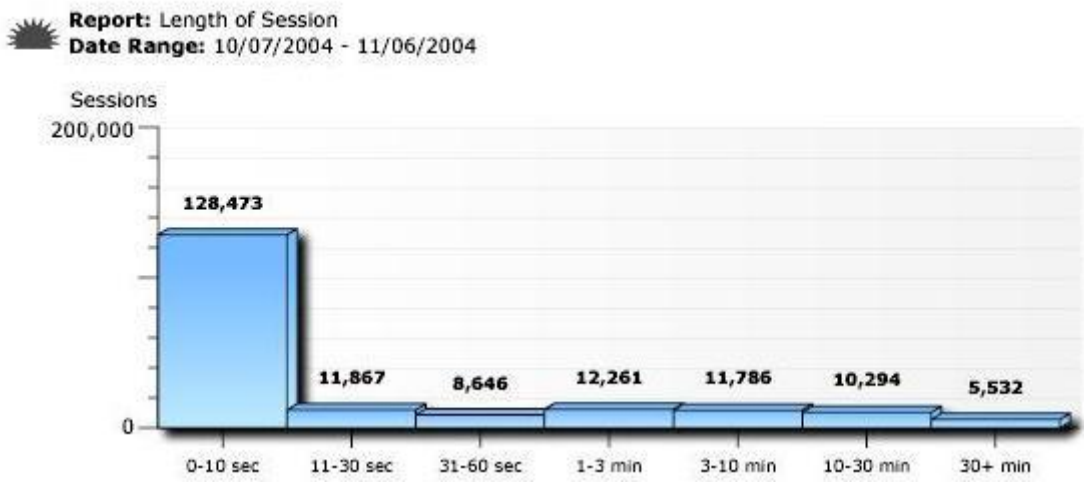


Fig. 14. Session lengths.

Surprisingly, the majority of sessions only last between zero to ten seconds, with ranges between one to three minutes and eleven to thirty seconds coming in second and third respectively. This is probably because most people on Iupac.org are finding what they are looking for fairly quickly, and then exiting the browser after either finding it, reading it or downloading it. This would make sense, since most respondents on the survey felt the navigation on the site was either easy or somewhere between easy and difficult.

Pages (1-10) / 20,917	Pageviews	Percent
1. /	13,168	1.71%
2. /dhtml_home.html	11,092	1.44%
3. /general/search.php	8,208	1.07%
4. /index_to.html	5,634	0.73%
5. /cgi-bin/htsearch	4,352	0.57%
6. /publications/compendium/index.html	3,072	0.40%
7. /publications/index.html	2,836	0.37%
8. /general/FAQs/ns.html	2,769	0.36%
9. /general/index.html	2,514	0.33%
10. /general/FAQs/elements.html	2,277	0.30%

Fig. 15. Pageviews.

Pageviews, which again is a request to the web server by a visitor's browser for a web page, averaged 24,839.35 a day, with the lowest Pageviews per day being 13,877 and the highest being 37, 218. The most popular pages were the main page to iupac.org (“/”, “/dhtml_home” – a DHTML menu driven version of the main page, and “/index_to.html” – a Text Only version of the site for browsers that do not support DHTML), the search engine (“/general/search.php”), the publications index and the FAQ for Iupac. Naturally it makes sense for the main page of a site to have the most page views, but it also makes sense that the publications index was in the top ten pages

viewed, since the majority of respondents in the web survey stated that they visited that section the most.

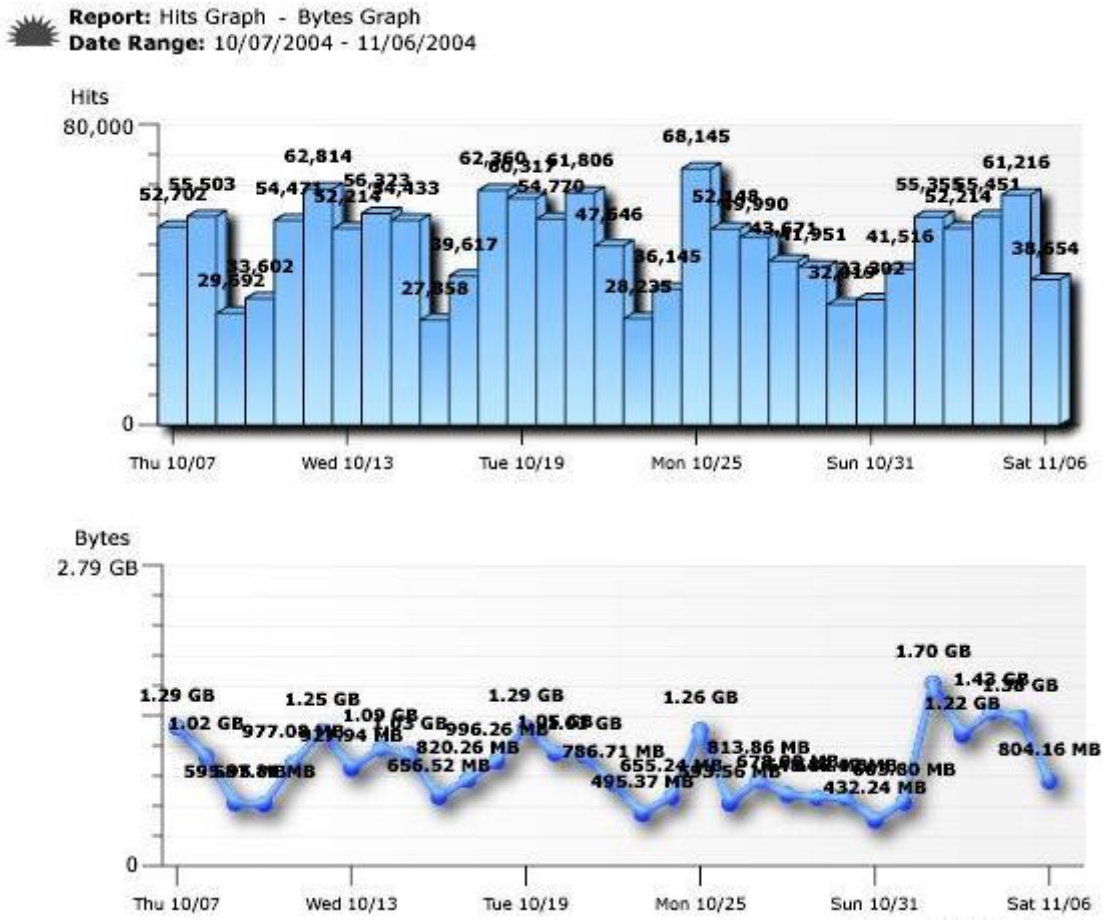


Fig. 16. Hits and bytes.

Iupac.org had a total of 1,149,140 hits, or successful requests to the web server from a visitor's browser for the month and averaged 48,262.58 hits per day. A total of 28.44 gigabytes, which is 28,440 megabytes, or 29,122,560 bytes were transferred for the month, averaging 939.58 megabytes a day.

File Types (1-10) / 37	Bytes	Percent	0 76.01%
1. pdf	21.62 GB	76.01%	
2. html	3.50 GB	12.31%	
3. jpg	1.89 GB	6.64%	
4. gif	667.98 MB	2.29%	
5. (no type)	235.44 MB	0.81%	
6. js	132.14 MB	0.45%	
7. ppt	92.99 MB	0.32%	
8. htm	83.85 MB	0.29%	
9. zip	69.22 MB	0.24%	
10. php	55.90 MB	0.19%	

Fig. 17. File types.

Taking a look at the breakdown of file types on Iupac.org, it is no surprise that the site averaged 939.58 megabytes transferred a day, considering that 76% of the files available are PDF files, at a total size of 21.62 gigabytes.

Total Sessions	188,859.00
Total Pageviews	770,020.00
Total Hits	1,496,140.00
Total Bytes Transferred	28.44 GB
Average Sessions Per Day	6,092.23
Average Pageviews Per Day	24,839.35
Average Hits Per Day	48,262.58
Average Bytes Transferred Per Day	939.58 MB
Average Pageviews Per Session	4.08
Average Hits Per Session	7.92
Average Bytes Per Session	157.93 KB
Average Length of Session	00:05:41

Fig. 18. Summary of sessions, pageviews, hits and bytes transferred.

Click Paths. Click Paths, which are the route that a visitor takes while visiting site after coming on to the main page, are displayed below in figure 30.

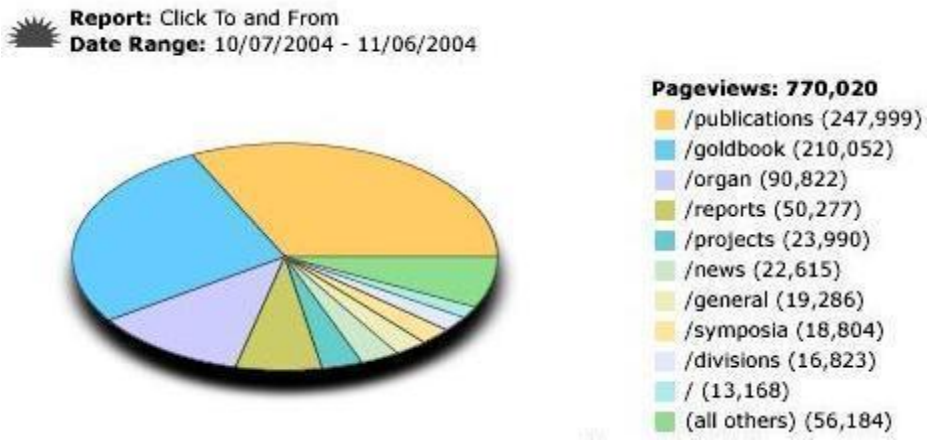
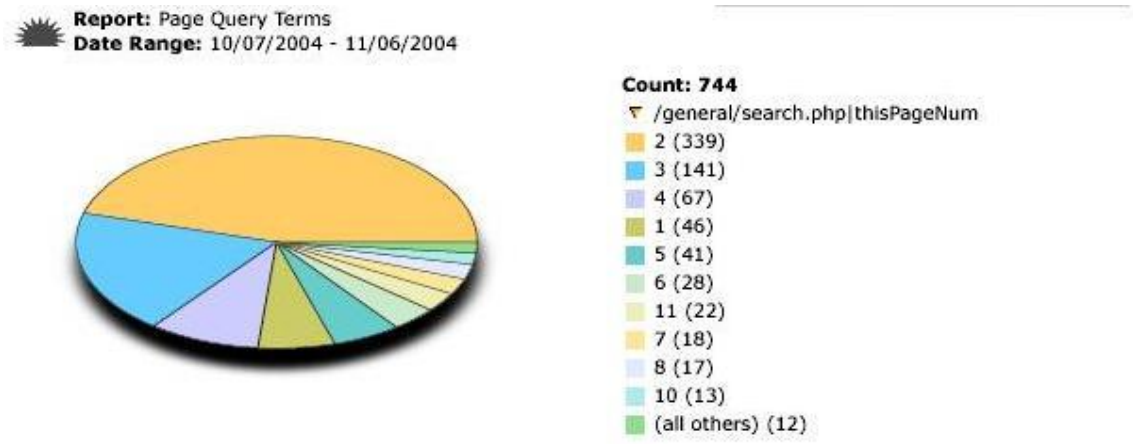


Fig. 19. Clicks to and from.

As expected, most users go to the publications section after coming into Iupac.org, which was also reflected in the web survey. The "/reports", "/goldbook" and "/symposia" paths are in the "Readers Corner" section of the site.



Page Query Terms (1-10) / 331	Count	Percent	0 4.17%
/general/search.php|query	744	100.00%	
arsenic	31	4.17%	
peak	12	1.61%	
nmr nomenclature	12	1.61%	
trace element reference value	12	1.61%	
chromatography	12	1.61%	
baseline	12	1.61%	
ultra	10	1.34%	
glossary and analytical	10	1.34%	
density metal table	10	1.34%	
exchange capacity q	8	1.08%	

Fig. 20. Query search terms.

Search Queries. Search queries, or the keywords used on the Iupac.org search, are displayed above in figure 20. Like the referral search terms visitors used to come to Iupac.org, it seems users use between two and three words in queries, and rarely use any boolean operators such as “or”, “and”, or “not”. As expected, although there were some queries that appeared to be popular for the month, like “arsenic”, the overall keywords in queries varied greatly. The pie chart at the top of figure 31 shows the number of hits for each “thisPageNum” link, which are the results pages links that are

displayed at the bottom of each search. From the results, it looks as if most people using the search were willing to search through two to three pages of results, something which is also reflected in the web survey. Unfortunately, given the way the search engine was implemented, there is no way to determine the number of successful searches versus the number of unsuccessful ones where no results are returned.

IX. Conclusion

What should be done to iupac.org in order to make the site better for its intended audience of chemistry enthusiasts and professionals? First, looking at what offered and visitor feedback, it seems clear to see that content does not have to be added or removed from the site. However, the organization of that content is something that will probably have to be addressed. Although the majority of participants felt that Iupac.org did not need any improvements, suggestions were primarily associated with the organization and navigation of the site. Among the suggestions were a better navigation menu, an easier to read publication list display, and journals in HTML format as opposed to PDF format. Given the large size of Iupac.org, it could certainly benefit from a better navigation system, possibly by adding a path or tree aspect to the navigation to let visitors see where they are on the site and where they can go.

On the subject of content, after looking at the areas that need improvement, it is important not forget about the areas that are the most popular. Both the survey and the web analytics showed that the most popular areas of the site are the main page, the search engine, the publications index and the FAQ (Frequently Asked Questions) section. With that in mind, it would be a good idea to do more than improve these areas, but rather to refine them to take full advantage of their popularity. The main page like many other pages on the site, would most benefit from a better navigation system, as well as more in-depth and descriptive labels. One participant stated that it took him a while to find what

he was looking for, and another one said that he had to search through a lot of the site to find things, since the navigation was not clear about where things are. The publications section also needs to be refined as it was ranked by Urchin as the number one destination that visitors of the site go to after entering the main page. One possible way to refine that section is to make aesthetic changes to the area, to make it more attractive to the large audience of people who already come there, and to possibly attract a new audience, as well as possibly offering different versions of the index and journal files themselves. The FAQ section could also be refined by essentially keeping it up to date and making frequent changes to it based on user feedback. Therefore, it would be a good idea to offer a separate feedback and comments page, as opposed to having it as part of the "Contact" section. One of the participants stated that he had trouble using the search engine, and felt that instructions should have been provided. That is something that could easily be remedied by offering a section about the search engine in the FAQ.

Finally, Iupac.org would be better for its intended audience of chemistry enthusiasts and professionals if the search engine, were refined. Search was one of the 4 most popular areas of the site, according to participants and Urchin. Participants of the survey wanted the search function to be able to search within more areas of the site, mainly the nomenclature and technology section. In addition to a set of instructions on how to use the search, participants also wanted the ability to search within the results of a previous search, to potentially provide an even higher level of search accuracy. However, it should be noted that the most participants did not have any problems with the search function, and web analytics of average searches done by visitors of the site suggested that most searches are performed without any Boolean conditionals ("and",

"not", "or", "+", and "~"). Given the way most search engines are designed, these conditionals can easily improve the accuracy of a search, and by not using them, users are limiting the potential accuracy of their searches. Although refinements to the search engine would be greatly appreciated, and extending the overall range of the search engine to more areas of the site would not be a problem, in order to improve the search engine, as mentioned earlier, it would be a good idea to offer a FAQ section on how to efficiently search to get the maximum performance out of the search engine. Both participants and Urchin confirmed that users on average would only search through one to three pages of search results. Given the fact that visitors on average were not performing efficient searches, the chances of finding what someone is looking for within those 3 results pages decreases significantly. Above all else, visitors of the site need to develop better search skills.

In conclusion, in order to make Iupac.org a better place for chemistry enthusiasts and professionals alike, changes need to be made to the overall navigation of the site, the organization of the publications section, popular areas of the site like the FAQ section, and the search engine itself. Iupac.org's goal is to become the definitive source for online digital chemistry resources, and by making these changes, they will be one step closer to reaching it.

Appendix A: Complete List of Urchin Features

Overall Features

- SVG-based Business Ready(sm) interactive charts and graphs - copy/paste into any Office application
- UTM first-party-cookie enhanced data collection (not available via ASP solutions) - no privacy problems or security warnings
- UTM supports cross-site tracking
- Pie chart, bar, or line-graph options for most reports
- P
-
- rint-friendly report-viewing mode
- Direct export to Word, Excel, and more
- Fully customizable reports - compare anything to anything
- Arbitrary Date Range analysis
- Visitors & Sessions reports not confined to monthly analysis (competitors are)
- Most accurate Unique Visitor reporting

Reporting Features

- Visitor Loyalty ("stickyness") report
- Help text embedded with calculation methodology explained
- All individual items in reports graphable over time
- E-commerce reporting - analyze your shopping cart vs. standard web traffic*
- Revenue Source reporting - geographical data on purchases*
- New Downloads report shows all files downloaded from your site (not available via ASP solutions)
- New Drilldown reports show information succinctly yet in complete detail
- Search Engine Marketing / SEO - Page Query Terms report automatically shows results of Cost-Per-Click campaigns, internal searches, and more
- Search Terms report shows actual keywords typed into search engines
- New IP Address and IP Drilldown reports
- Intranet IP analysis
- New Robots and Spiders reports (not available via ASP solutions)
- New Client Parameters reports - screen colors, resolution, timezone offset, Java/Javascript versions, etc.
- Automatic exclusion of "bot" traffic from Visitors reports
- Click-path analysis improved (click to/from report)
- Improved flexibility in Usernames reporting - can be parsed out of any log field, and can be used for visitor identification and session tracking

Admin Features

- Remote administration and report-viewing built-in
- All-Profiles Report ranks traffic to all sites analyzed
- Wizard-based setup for all major features
- Region settings configurable per-user (time / date formats, etc.)
- Reporting language default configurable per-user
- Help text embedded
- Users & Groups management / authentication built-in
- Multiple Admin levels for hosting resellers
- Configurable run priority, memory and processor usage, and database size
- Built-in Scheduler to manage Urchin jobs
- Automatic log format detection
- Automatic archiving of past months' databases to reduce storage overhead
- FTP log retrieval now supports wildcard (POSIX) expressions - easily grab remote log files
- UNC paths to logfiles now supported
- Pre-built Filters
- Custom log formats supported - analyze logs from any web server
- IIS configuration export utility included to ease administration for IIS hosts with thousands of sites
- Include/exclude filters now available for all Admin screens - useful for installations with large numbers of sites

Hosting-Specific Features

- Portal integration
- User-authentication bypass option
- Customizable report sets for different hosting levels
- Direct linking to specific reports

System Features

- DNS Database built-in for incredible lookup performance - auto updates
- Speed of log processing improved approximately 30%
- Fault tolerance/recovery built-in
- Automatic database backups and recovery
- Automatic archiving of past months' databases to reduce storage overhead
- Fully scriptable operation

Licensing Features

- Modular Licensing - only pay for the advanced features you need
- E-commerce Module - analyze shopping cart data
- Load-balancing Module - analyze sites using more than one server
- 100 Profile Pack Module - add capacity to your Urchin installation
- Campaign Tracking Module
- Geotargeting Module - *coming soon!*

*=Requires additional Module at extra cost

Report List

- Traffic

    - Sessions Graph
    - Pageviews Graph
    - Hits Graph
    - Bytes Graph
    - Summary
    - Load Balancing
        - Log Source by Hits
        - Log Source by Bytes
- Visitors & Sessions
    - Visitors by Day
    - Sessions by Day
    - Unique Visitors
    - Unique Sessions
    - Visitor Loyalty
    - Session Frequency
    - Summary
- Pages & Files
    - Requested Pages
    - Downloads
    - Page Query Terms
    - Posted Forms
    - Status and Errors
    - All Files
        - All Files by Hits
        - All Files by Bytes
    - Directory Drilldown
        - Directory by Pages Drilldown
        - Directory by Files Drilldown
        - Directory by Bytes Drilldown
    - File Types
        - File Types by Hits
        - File Types by Bytes
- Navigation
    - Entrance Pages
    - Exit Pages
    - Click Paths
    - Click To and From
    - Length of Pageview
    - Depth of Session
    - Length of Session
- Referrals
    - Referrals
    - Referral Drilldown
    - Search Terms
    - Search Engines
    - Referral Errors
- Domains & Users
    - Domains
    - Domain Drilldown
    - Countries
    - IP Addresses
    - IP Drilldown
    - Usernames
        - Usernames by Hits
        - Usernames by Bytes
        - Usernames by Sessions

- Browsers & Robots
    - Browsers
        - Browsers by Sessions Drilldown
        - Browsers by Hits Drilldown
        - Browsers by Bytes Drilldown
    - Platforms
        - Platforms by Sessions Drilldown
        - Platforms by Hits Drilldown
        - Platforms by Bytes Drilldown
    - Combos
        - Combos by Sessions
        - Combos by Hits
        - Combos by Bytes
    - Robots
        - Robots by Hits Drilldown
        - Robots by Bytes Drilldown
- Client Parameters
    - Screen Resolution
    - Screen Colors
    - Languages
    - Java Enabled
    - Timezone Offset
    - Javascript Version

E-Commerce Module

- E-Commerce
    - Revenue
    - Number of Transactions
    - E-Commerce Summary
    - Products
        - Products by Revenue
        - Products by Quantity
        - Products by Revenue Drilldown
        - Products by Quantity Drilldown
    - Revenue Source
        - Revenue by Region Drilldown
        - Revenue by City
        - Revenue by Referrals
        - Revenue by Search Terms
        - Revenue by Search Engines Drilldown
        - Revenue by Domains Drilldown

Campaign Tracking Module

- Campaign Tracking
    - Overall Results
    - Goal Results
    - Lead Sources
        - Acquisition

        - Quality
        - Conversion
        - ROI
    - Keyword Analysis
        - Acquisition
        - Quality
        - Conversion
        - ROI
    - Keyword Comparison
        - Acquisition
        - Quality
        - Conversion
        - ROI
    - Campaign Comparison
        - Acquisition
        - Quality
        - Conversion
        - ROI
    - Medium Comparison
        - Acquisition
        - Quality
        - Conversion
        - ROI
    - Content (A/B) Testing
        - Acquisition
        - Quality
        - Conversion
        - ROI
    - Latency Reports
        - Time to Goal
        - Sessions to Goal
        - Time to Transaction
        - Sessions to Transaction
    - Day Parts Breakdown
        - Goal Conversion by Hour
        - Sales Conversion by Hour
    - Click Fraud Watch
        - Repeat Clicks by IP
        - Repeat Clicks by Source

Bibliography

1. International Union of Pure and Applied Chemistry. 2004. 21 July 2004 <http://www.iupac.org>

2. Jones, Steve; Cunningham, Sally Jo; McNab, Roger; Boddie, Stefan; "A Transaction Log Analysis of a Digital Library" International Journal on Digital Libraries. 1999. 04 August 2004. < http://www.cs.waikato.ac.nz/~stevej/Research/PAPERS/ijodllogs.pdf>

3. D'Alessandro, Michael P.; D'Alessandro, Donna M.; Galvin, Jeffery R.; Erkonen, William E.; "Evaluating Overall Usage of A Digital Health Sciences Library" Journal of the American Medical Library Association. 2003. 06 August 2004. < http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=226457&action=stream&blobtype=pdf>

4. Vakkayil, Jacob D. "Digital Libraries – Some Analog Issues" University of Arizona. 2004. 20 August 2004. < http://dlist.sir.arizona.edu/archive/00000306/02/digitlib.pdf>

5. Harter, Steven P. "Scholarly Communication and the Digital Library: Problems and Issues" Journal of Digital Information. 1997. 22 August 2004. < http://jodi.ecs.soton.ac.uk/Articles/v01/i01/Harter/>

6. Bullock, Alison. "Preservation of Digital Information: Issues and Current Status" Network Notes 60 (Apr. 1999): 29 August 2004. < http://www.collectionscanada.ca/9/1/p1-259-e.html>

7. Xie, Hong, and Dietmar Wolfram. "State Digital Library Usability: Contributing Organizational Factors" Journal of Digital Information. 2002. 02 September 2004. < http://portal.acm.org/citation.cfm?id=772460>

8. Ke, Hao-Ren; Kwakkelaar, Rolf; Tai, Yu-Min; Chen, Li-Chun; "Exploring Behavior of E-journal Users in Science and Technology: Transaction Log Analysis of Elsevier's ScienceDirect OnSite in Taiwan" Library & Information Science Research. 2002. 04 September 2004. < http://portal.acm.org/citation.cfm?id=985363.985368&coll=GUIDE&dl=ACM>

9. Bishop, Ann P. "Digital Library Use: Social Practice in Design and Evaluation" Journal of the ACM. 2003. 10 September 2004. < http://portal.acm.org/ft_gateway.cfm?id=336700&type=pdf>

10. Cullen, Rowena. "Evaluating Digital Libraries in the Health Sector. Part 1: Measuring Inputs and Outputs" Health Information and Libraries Journal. 2003. 17 September 2004. <http://www.blackwellpublishing.com/issue.asp?iid=4&ref=1471-1834&vid=20>

11. Cullen, Rowena. "Evaluating Digital Libraries in the Health Sector. Part 2: Measuring Impacts and Outcomes" Health Information and Libraries Journal. 2003. 17 September 2004. < http://www.blackwellpublishing.com/issue.asp?iid=4&ref=1471-1834&vid=20>

12. Greenstein, Daniel; Thorin, Suzanne E; "The Digital Library: A Biography" Digital Library Foundation. 2002. 24 September 2004. < http://www.clir.org/pubs/reports/pub109/pub109.pdf>

13. Columbia University Library Section Criteria for Digital Imaging 2001. 24 September 2004. <http://www.columbia.edu/cu/libraries/digital/criteria.html>

14. Grothkopf, Uta. "Bits and Bytes and Still a Lot of Paper: Astronomy Libraries and Librarians in the Age of Electronic Publishing" Astrophysics and Space Science. 247 (1997): 155-174

15. Dhyani, D., Ng W., & Bhowmick S; "A survey of Web metrics." ACM Computing Surveys, 34 (4), 469-503. 2002. 01 October 2004 < portal.acm.org/ft_gateway.cfm?id=592645&type=pdf>

16. "E-metrics: Business Metrics for the New Economy." Netgen.com. 2003. 05 October 2004. <http://netgen.com/emetrics/emetrics.pdf>

17. Smith, Arthur P. "Web Performance Metrics for online Journals: Monitoring & Improving Accessibility." The American Physical Society

18. Regan, Keith. "Web Performance Metrics That Matter." Ecommercetimes.com. 2003. 05 October 2004. <http://www.ecommercetimes.com/perl/story/31480.html>

19. Patton, Susannah. "Web Metrics That Matter." Cio.com. 2003. 06 October 2004. <http://www.cio.com/archive/111502/matter.html>

20. Urchin Web Analytics Software. 2004. 20 October 2004. <http://www.Urchin.com>

21. phpESP – php Easy Survey Package 2004. 20 October 2004 <http://phpesp.sourceforge.net/>